

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE CIENCIAS BIOLÓGICAS**  
**Departamento de Bioquímica y Biología Molecular I**



**CARACTERIZACIÓN ESTRUCTURAL MEDIANTE  
MICROSCOPIA ELECTRÓNICA DEL COMPLEJO  
Xrcc4-DNA Ligasa IV, IMPLICADO EN  
REPARACIÓN DE ROTURAS DEL DNA.**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR  
PRESENTADA POR**

**María Ángeles Recuero Checa**

Bajo la dirección del doctor

Óscar Antonio Llorca Blanco

**Madrid, 2011**

**ISBN: 978-84-694-3762-9**

**© María Ángeles Recuero Checa, 2010**

# Caracterización estructural mediante microscopía electrónica del complejo Xrcc4-DNA Ligasa IV, implicado en reparación de roturas del DNA

María Ángeles Recuero Checa

Tesis Doctoral

Madrid, 2010



Universidad Complutense de Madrid  
Facultad de Ciencias Biológicas  
Departamento de Bioquímica y Biología molecular I



Directorio

---

Índice general

Introducción

Objetivos

Materiales y Métodos

Resultados

Discusión

Conclusiones

Bibliografía

Publicaciones





**UNIVERSIDAD COMPLUTENSE DE MADRID**

Facultad de Ciencias Biológicas

Departamento de Bioquímica y Biología Molecular I



**Caracterización estructural mediante microscopía electrónica  
del complejo Xrcc4-Dna Ligasa IV, implicado en reparación de  
roturas del DNA**

TESIS DOCTORAL

María Ángeles Recuero Checa

Madrid, 2010







**UNIVERSIDAD COMPLUTENSE DE MADRID**

Facultad de Ciencias Biológicas

Departamento de Bioquímica y Biología Molecular I



**Caracterización estructural mediante microscopía electrónica  
del complejo Xrcc4-Dna Ligasa IV, implicado en reparación de  
roturas del DNA**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR PRESENTADA  
POR**

**María Ángeles Recuero Checa**

Esta Tesis ha sido realizada en el Centro de Investigaciones Biológicas (CSIC) bajo la dirección del Dr. Óscar Antonio Llorca Blanco y financiado por el programa de Formación de Personal Investigador del Ministerio de Ciencia e Innovación (BES-2006-13657)

Madrid, 2010







***A mis padres***





*"Después de escalar una montaña muy alta,  
descubrimos que hay muchas otras montañas por escalar"*  
*(Nelson Mandela)*





# Agradecimientos

Después de casi seis años en el CIB resulta difícil resumir en unas pocas palabras todas las cosas que he vivido durante este tiempo y la gente con la que lo he disfrutado.

En primer lugar tengo que agradecer al Ministerio de Ciencia e Innovación y al programa de becas FPI, por haberme financiado durante la mayor parte de este tiempo. También tengo que dar las gracias al grupo de Laurence Pearl en Londres, especialmente a Andy por la purificación de las proteínas con las que he trabajado.

Volviendo al CIB, tengo que agradecerte en primer lugar, Óscar, el haberme dado la oportunidad de trabajar en tu laboratorio. De tí me quedo con tu infinita paciencia (creo que lo hemos mencionado todos en los agradecimientos, pero es que es cierto), la preocupación que siempre tienes por la gente que está a tu cargo, y tu optimismo y energía frente a cualquier problema. A pesar de toda la vorágine que te rodea, siempre sacas tiempo para atendernos a todos. Mil gracias por todo eso.

También me gustaría dar las gracias a toda la gente del labo, los que ya no están aquí y con los que he pasado buenísimos momentos: Jaska, por seguir siendo una buena amiga; Angelito, nunca olvidaré el día en el que se “disolvió” mi gel; Rafa, a pesar de su “carácter”; Isra, el gran mandril. A los que todavía siguen aquí: Eva y Bego, muchas gracias por vuestra ayuda en el día a día del cacharreo, y los consejos que me habéis dado. A los que acaban de llegar hace poco, porque también hemos pasado buenos ratos: Andrés, Martín, César y Roberto, os deseo mucha suerte. A Ernesto, por tantas cosas que podrían ocupar perfectamente una tesis. Tanto profesional como personalmente te he admirado desde que te conocí, y con el tiempo has pasado a ser esencial en mi vida. Gracias por todo el tiempo que me dedicas.

Quiero dar las gracias también a Lola Guirao y Carmen Terrón, que también se han marchado del CIB y a las que recuerdo con mucho cariño.

No me puedo olvidar de muchos amigos que están todavía aquí. Luque, te estaré eternamente agradecida por recomendarme las croquetas del Melo's; Elenita, espero veros pronto a tí y a Gonzalo. A María, por tantos momentos

compartidos dentro y fuera del labo. También a Nacho y a Mon, porque los viajes con vosotros tienen sorpresa asegurada y muchas risas.

Me gustaría dar las gracias también a los microscopistas del CNB, con los que seguimos compartiendo buenos momentos. Y del CNB también a Patricia, has sido todo un descubrimiento tanto científico como personal.

A Eva Nogales y la gente de su labo de Berkeley, os quiero dar las gracias por tratarme como una más durante mi estancia allí, especialmente a Jordi, Claudio y Aurélie. Os volveré a ver muy pronto.

A mis amigas, Coke, Mariájo y Menchu. No tengo palabras para describir lo que significáis para mí. Gracias por estar siempre ahí.

Tengo que agradecer también a toda la gente que ha pasado por mi vida y que ha dejado su pequeña huella en ella, aunque por unas u otras razones hayamos terminado distanciándonos.

Quiero agradecer también a Álvaro y Mari Luz, por tratarme como una hija más.

Por último, me gustaría dar las gracias a toda mi familia, y sobre todo, a mis padres, Ángel y Paquí, por todo el esfuerzo que han hecho para que esté aquí y por su cariño. Gracias a mi hermana Loly y a Jose, y al último en llegar a la familia, Rubén, la alegría de mi casa.







# Índice general

<b>1. Introducción .....</b>	<b>1</b>
1.1. Las roturas en la doble cadena de DNA son esenciales en determinados procesos celulares .....	3
1.2. La activación de los mecanismos de señalización celular ocurre en respuesta al daño en el DNA.....	5
1.3. La reparación de las roturas en la doble cadena de DNA es llevada a cabo mediante dos mecanismos diferentes.....	5
1.3.1. El sistema de reparación basado en recombinación homóloga.....	6
1.3.2. El mecanismo de reparación mediante unión de extremos no homólogos ( <i>Non-Homologous End-Joining</i> o NHEJ) .....	7
1.4. El proceso de ligación durante la reparación mediante NHEJ es llevado a cabo por el complejo Xrcc4-DNA ligasa IV .....	10
1.5. Estructura de la DNA ligasa IV, la ligasa específica del proceso de reparación mediante unión de extremos no homólogos .....	11
1.6. Características estructurales de Xrcc4 .....	15
1.7. XLF/Cernunnos interviene en determinados procesos de reparación mediante NHEJ .....	16
1.8. La microscopía electrónica aplicada a la reconstrucción tridimensional de partículas individuales .....	19
1.8.1. Perspectiva de la microscopía electrónica de partículas individuales.....	19
1.8.2. Introducción a la metodología empleada en la obtención de una estructura tridimensional mediante microscopía electrónica .....	21
<b>2. Objetivos .....</b>	<b>27</b>
<b>3. Materiales y Métodos.....</b>	<b>31</b>
3.1. Clonación, expresión y purificación de proteínas .....	33
3.1.1. Clonación, expresión y purificación de Xrcc4 de <i>Mus musculus</i> .....	33
3.1.2. Clonación, expresión y purificación del complejo Xrcc4-DNA ligasa IV humano.....	33

3.1.3. Clonación, expresión y purificación del complejo Xrcc4( $\Delta$ 214-336)-DNA ligasa IV .....	34
3.1.4. Clonación, expresión y purificación de XLF humana, XLF( $\Delta$ 242-299) y XLF( $\Delta$ 248-299) .....	34
3.1.5. Purificación de Xrcc4 mediante centrifugación preparativa en gradientes de glicerol .....	35
3.1.6. Purificación del complejo Xrcc4-DNA ligasa IV mediante cromatografías de exclusión molecular .....	35
3.2. Electroforesis en condiciones nativas y desnaturizantes (SDS-PAGE) .....	35
3.3. Experimentos de ultracentrifugación analítica .....	36
3.4. Ensayos de entrecruzamiento con glutaraldehído .....	37
3.5. Caracterización de los complejos entrecruzados mediante espectrometría de masas .....	38
3.6. Ensayos de retardo en la movilidad electroforética en complejos proteína-DNA .....	38
3.7. Ensayos de interacción entre el complejo Xrcc4-DNA ligasa IV y la proteína XLF .....	40
3.8. Análisis estructural mediante microscopía electrónica y procesamiento de imágenes .....	40
3.8.1. Preparación de muestras mediante tinción negativa .....	40
3.8.2. Adquisición y digitalizado de imágenes .....	40
3.8.3. Procesamiento de imágenes y reconstrucción tridimensional .....	40
3.8.4. Cálculo de la resolución y visualizado de los volúmenes .....	41
3.9. Predicción de estructura secundaria y alineamientos de secuencias .....	42
3.10. Modelado de proteínas y ajuste en la reconstrucción tridimensional obtenida mediante microscopía electrónica .....	42
3.11. Marcaje de proteínas mediante moléculas de oro unidas a níquel .....	43
3.12. GraFix .....	44
<b>4. Resultados .....</b>	<b>47</b>
4.1. Purificación de Xrcc4 de ratón .....	49

4.2. Ensayos de entrecruzamiento de Xrcc4 con glutaraldehído .....	50
4.3. Análisis de Xrcc4 mediante ultracentrifugación analítica .....	51
4.4. Purificación de los complejos diméricos de Xrcc4 mediante gradientes de glicerol previo a su análisis mediante microscopía electrónica .....	52
4.5. Análisis de la interacción de Xrcc4 con DNA .....	52
4.6. Análisis estructural de Xrcc4 de ratón .....	54
4.6.1. Reconstrucción tridimensional mediante microscopía electrónica .....	54
4.6.2. Modelado del dominio C-terminal de Xrcc4 .....	55
4.6.3. Encaje de los modelos generados por Robetta en el mapa de densidad electrónica.....	56
4.7. Purificación y caracterización del complejo Xrcc4-DNA ligasa IV humano .....	58
4.8. Experimentos de entrecruzamiento del complejo Xrcc4-DNA ligasa IV mediante glutaraldehído .....	59
4.9. Análisis de la interacción del complejo Xrcc4-DNA ligasa IV con DNA .....	60
4.10. Análisis estructural del complejo Xrcc4-DNA ligasa IV mediante microscopía electrónica .....	61
4.11. Localización del extremo C-terminal de Xrcc4 en el complejo Xrcc4- DNA ligasa IV.....	62
4.12. Localización del extremo N-terminal de la DNA ligasa IV en el complejo Xrcc4-DNA ligasa IV mediante marcaje con moléculas de oro .....	64
4.13. Reconstrucción tridimensional del complejo Xrcc4-DNA ligasa IV .....	66
4.14. Purificación del complejo Xrcc4-DNA ligasa IV estabilizado mediante glutaraldehído (GraFix) .....	70
4.15. Purificación de XLF humana.....	71
4.16. Análisis de la interacción XLF-DNA mediante geles nativos.....	74
4.17. Caracterización de la interacción entre XLF y el complejo Xrcc4- DNA ligasa IV.....	75
<b>5. Discusión .....</b>	<b>79</b>
5.1. La estructura tridimensional de Xrcc4 completa revela una región	

carboxilo-terminal globular .....	82
5.2. La región N-terminal de la proteína DNA-ligasa IV se localiza próxima a la región N-terminal de Xrcc4 en el complejo, mientras que la región C-terminal se localiza en el extremo opuesto.....	87
5.3. La estructura tridimensional del complejo Xrcc4-DNA ligasa IV es consistente con los datos obtenidos mediante el análisis bidimensional previo.....	89
5.4. Modelo estructural del complejo Xrcc4-DNA ligasa IV e implicaciones fisiológicas .....	90
5.5. Xrcc4 y DNA ligasa IV sólo interaccionan de forma estable con DNA en presencia del heterodímero Ku. ....	92
5.6. La región C-terminal de la proteína XLF introduce inestabilidad que dificulta la purificación de la proteína completa .....	93
5.7. El proceso de ligación en el contexto del NHEJ.....	94
<b>6. Conclusiones .....</b>	<b>99</b>
<b>7. Bibliografía .....</b>	<b>103</b>
<b>8. Publicaciones.....</b>	<b>119</b>

## Abreviaturas y símbolos

2D	Bidimensional
3D	Tridimensional
A	Absorbancia
Å	Angstrom
ADP	Adenosín 5'-difosfato (del inglés <i>Adenosine diphosphate</i> )
AMP	Adenosín 5'-monofosfato (del inglés <i>Adenosine monophosphate</i> )
ATM	Del inglés <i>Ataxia Telangiectasia mutated</i>
ATP	Adenosín 5'-trifosfato (del inglés <i>Adenosine triphosphate</i> )
ATR	Del inglés <i>ATM and Rad3 related</i>
BRCT	Del inglés <i>Breast Cancer-1 (BRCA1) gene carboxy terminal</i>
BSOFT	Del inglés <i>Bernard's software</i>
CCD	Del inglés <i>Charged Coupled Device</i>
C-terminal	Extremo carboxilo terminal
CTF	Función de transferencia de contraste (del inglés <i>Contrast Transfer Function</i> )
DBD	Dominio de unión a DNA (del inglés <i>DNA-binding domain</i> )
DNA	Ácido desoxirribonucleico (del inglés <i>desoxirribonucleic acid</i> )
DNA-PKcs	Subunidad catalítica de la proteína quinasa dependiente de DNA (del inglés <i>DNA-dependent protein kinase catalytic subunit</i> )
DTT	Ditiotreitol
e <sup>-</sup>	Electrones
EDTA	Ácido etilendiaminotetraacético (del inglés <i>Ethylendiaminetetraacetic acid</i> )
EGTA	Ácido etilenglicoltetraacético (del inglés <i>Ethylenglycoltetraacetic acid</i> )
EMAN	Programa de análisis de micrografías electrónicas (del inglés <i>Electron micrograph analysis</i> )
EMSA	Ensayo de movilidad electroforética (del inglés <i>Electrophoretic mobility shift assay</i> )
FSC	Coeficiente de correlación en el espacio de Fourier (del inglés <i>Fourier shell correlation</i> )
GA	Glutaraldehído
GST	glutathione-S-transferasa
HCl	Ácido clorhídrico
HEPES	Ácido N-2-hidroxietilpiperacina-N'-2-etanosulfónico
HPLC	Cromatografía líquida de alta presión (del inglés <i>High performance liquid chromatography</i> )
IPTG	Isopropil-Beta-D-tiogalactopiranosido
KCl	Cloruro potásico
MALDI-TOF	Desorción/Ionización mediante láser asistida por matriz con tiempo de vuelo
mAu	Del inglés <i>miliAbsorbance units</i>
ML2D	del inglés, <i>Maximum-likelihood alignment in 2D</i>
ML3D	del inglés, <i>Maximum-likelihood alignment in 3D</i>
NAD	Nicotinamida-adenina dinucleótido
NTasas	Nucleotidiltransferasa

NHEJ	Unión de extremos no homólogos (del inglés <i>Non-homologous end-joining</i> )
Ni-NTA	Ácido nitrilotriacético acoplado a níquel (del inglés <i>Nickel-nitrilotriacetic acid</i> )
nt	Nucleótido
N-terminal	Extremo amino terminal
(OB)-fold	Plegamiento de unión a oligosacáridos y oligonucleótidos, del inglés <i>Oligomer binding-fold</i>
OD	Densidad óptica (del inglés <i>optical density</i> )
P	Fosfato
pb	Pares de bases
PAGE	Electroforesis en gel de poliacrilamida (del inglés <i>Polyacrylamide-gel electrophoresis</i> )
PBS	Tampón fosfato salino, del inglés <i>phosphate buffered saline</i>
PCR	Reacción en cadena de la polimerasa (del inglés <i>polymerase chain reaction</i> )
PDB	Banco de datos de proteínas, del inglés <i>Protein Data Bank</i>
Phyre	del inglés <i>Protein homology/analogy recognition Enghien</i>
PI3K	del inglés <i>phosphatidylinositol-3kinase</i>
PIKKs	del inglés <i>phosphatidylinositol-3-kinase related protein kinases</i>
PMSF	Fluoruro de fenilmetilsulfonilo (del inglés <i>phenylmethylsulfonyl fluoride</i> )
PNK	del inglés <i>polynucleotide kinase</i>
PSA	Persulfato amónico
p/v	Peso/volumen
PVDF	del inglés <i>polyvinylidene difluoride</i>
RAG	del inglés <i>recombinase-activating genes</i>
RMN	Resonancia magnética nuclear
rpm	Revoluciones por minuto
SAXS	Dispersión de rayos X de ángulo pequeño (del inglés, <i>Small Angle X-ray Scattering</i> )
SDS	Dodecil sulfato sódico, del inglés <i>sodium dodecylsulphate</i>
SPIDER	del inglés <i>system for processing image data in electron microscopy and related fields</i>
TBE	Tris-Borato EDTA
TEM	Microscopía de transmisión de electrones ( <i>Transmission electron microscopy</i> )
TEMED	N,N,N',N'-tetrametiletlen-diamina
Tris	Tris (hidroximetil) aminometano
v/v	Volumen/volumen
XLF	Del inglés <i>Xrcc4-like factor</i>
XMIPP	del inglés <i>X-Window-based microscopy image processing Packaged</i>
XRCC4	del inglés <i>X-ray cross-complementing defective repair in Chinese hamster cells 4</i>







# INTRODUCCIÓN

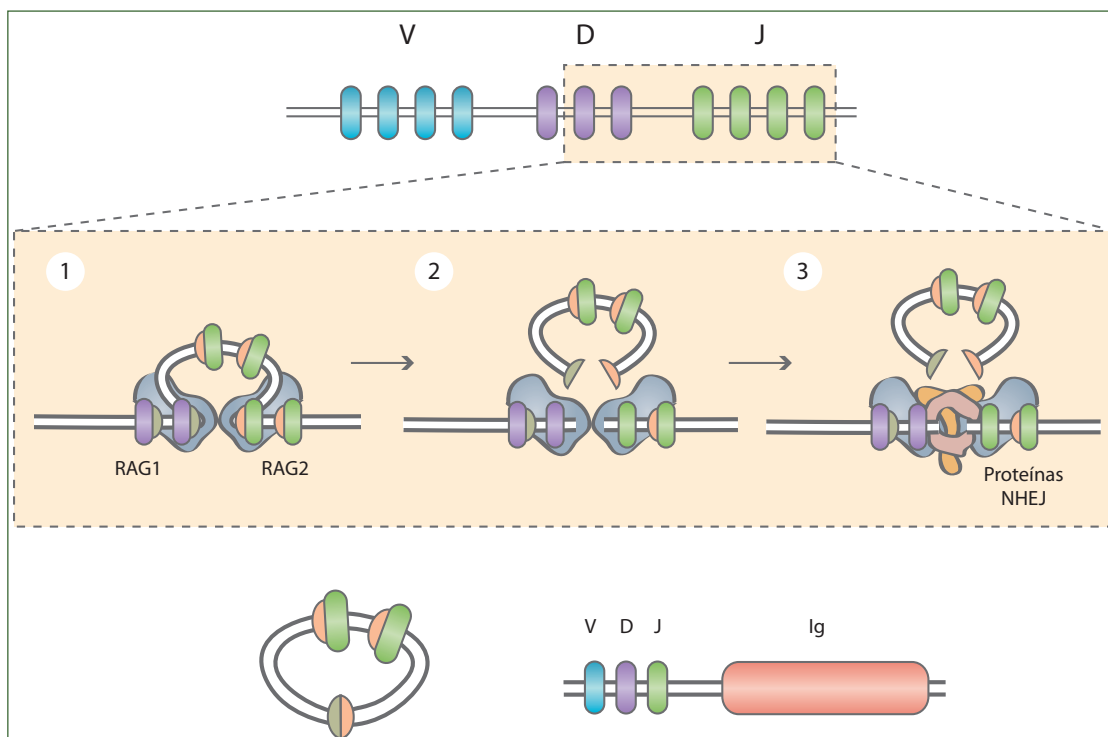


La integridad del material genético de las células de cualquier organismo vivo está sometida continuamente a multitud de amenazas, tanto externas como endógenas. Para proteger y preservar la información genética, así como para prevenir la transmisión de aberraciones cromosómicas a la descendencia, los diversos tipos de alteraciones estructurales en el DNA deben ser detectados y reparados durante todas y cada una de las etapas del ciclo celular (Burma et al., 2006; Hoeijmakers, 2001). Para llevarlo a cabo, las células eucariotas han desarrollado un conjunto de complejos mecanismos que integran la detección del daño en el DNA con la activación de un proceso de señalización que produce la detención del ciclo celular en los puntos llamados de control, con el fin de coordinar el proceso de reparación. Si éste se lleva a cabo con éxito el ciclo celular continúa, en caso contrario, se induce la apoptosis o muerte celular programada (Harrison y Haber, 2006; Zhou y Elledge, 2000). Todos estos procesos son llevados a cabo en las células por grandes complejos macromoleculares, los cuales se ensamblan de una forma dinámica en el lugar de la lesión.

Existen diversos tipos de roturas en el DNA, resultando las roturas de doble cadena uno de los más peligrosos para la supervivencia celular. Esto se debe principalmente a que si estas roturas no son reparadas correctamente y de una forma rápida, por ejemplo, si existe alguna mutación en proteínas de reparación o de control de la proliferación celular, se pueden producir aberraciones cromosómicas graves e inestabilidad genómica que pueden desembocar en el desarrollo de un proceso cancerígeno (Lieber et al., 2006; van Attikum y Gasser, 2009). Las mutaciones en proteínas implicadas en rutas de reparación también pueden provocar inmunodeficiencias y enfermedades neurodegenerativas, entre otras patologías, además de generar fenotipos celulares sensibles a la radiación. Estas roturas son reparadas en células eucariotas principalmente mediante dos mecanismos, dependiendo del momento del ciclo celular en el que tenga lugar la rotura, denominados recombinación homóloga y unión de extremos no homólogos (*Non-homologous end-joining* o NHEJ) (Burma et al., 2006; Povirk, 2006). Durante el desarrollo de esta tesis se han estudiado el complejo de proteínas Xrcc4-DNA ligasa IV, así como la proteína XLF que participan en la ruta de reparación de DNA de unión de extremos no homólogos.

### **1.1. Las roturas en la doble cadena de DNA son esenciales en determinados procesos celulares**

Las roturas en la doble cadena de DNA se caracterizan por estar espaciadas entre sí de 10 a 20 nucleótidos a lo largo de la cadena de DNA y suelen tener extremos 3' y 5' complementarios, si bien los extremos generados tras la rotura pueden ser de distintos tipos (Mahaney et al., 2009). Estas lesiones son producidas por diversos agentes externos, siendo los más destacados la radiación ionizante (rayos X y rayos  $\gamma$ ), así como determinados fármacos utilizados en quimioterapia (van Gent et al., 2001). Sin em-



**Figura 1.1. Esquema del proceso de recombinación V(D)J para generar la variabilidad en la respuesta inmune.** En este proceso participan determinadas proteínas del proceso de reparación NHEJ, entre ellas, el complejo Xrcc4-DNA ligasa IV y la proteína Artemis, esencial durante la recombinación V(D)J.

bargo, también existen procesos celulares que pueden provocar este tipo de roturas, como por ejemplo, el proceso de recombinación V(D)J que genera la variabilidad en los anticuerpos y los receptores de antígeno en las células del sistema inmune. El proceso de recombinación V(D)J consiste en una reorganización del DNA en regiones específicas del genoma durante el desarrollo de los linfocitos T y B para generar la diversidad que caracteriza a la respuesta inmune (Figura 1.1) (Sekiguchi y Ferguson, 2006). En este proceso, las nucleasas RAG1 y RAG2 provocan roturas intencionadas en regiones concretas de la doble cadena de DNA, conocidas como secuencias de recombinación RSSs (**R**ecombination **S**ignal **S**equences) adyacentes a regiones de DNA codificantes (Figura 1.1, panel 1). A continuación, se aísla el fragmento de DNA localizado entre los dos nuevos extremos generados (Figura 1.1, panel 2) y los fragmentos codificantes son ligados reconstituyendo la doble cadena de DNA (Figura 1.1, panel 3). Este proceso es llevado a cabo por algunas de las proteínas implicadas en el proceso de reparación mediante extremos no homólogos (NHEJ), como DNA-PKcs, Artemis (una proteína esencial en este proceso), Ku70/Ku80, Xrcc4 y DNA ligasa IV (Weterings y Chen, 2008). Se ha descrito que RAG2 es fosforilada específicamente por DNA-PKcs en un residuo serina, y esta fosforilación parece ser importante en la regulación del proceso de unión de los extremos del DNA (Hah et al., 2007).

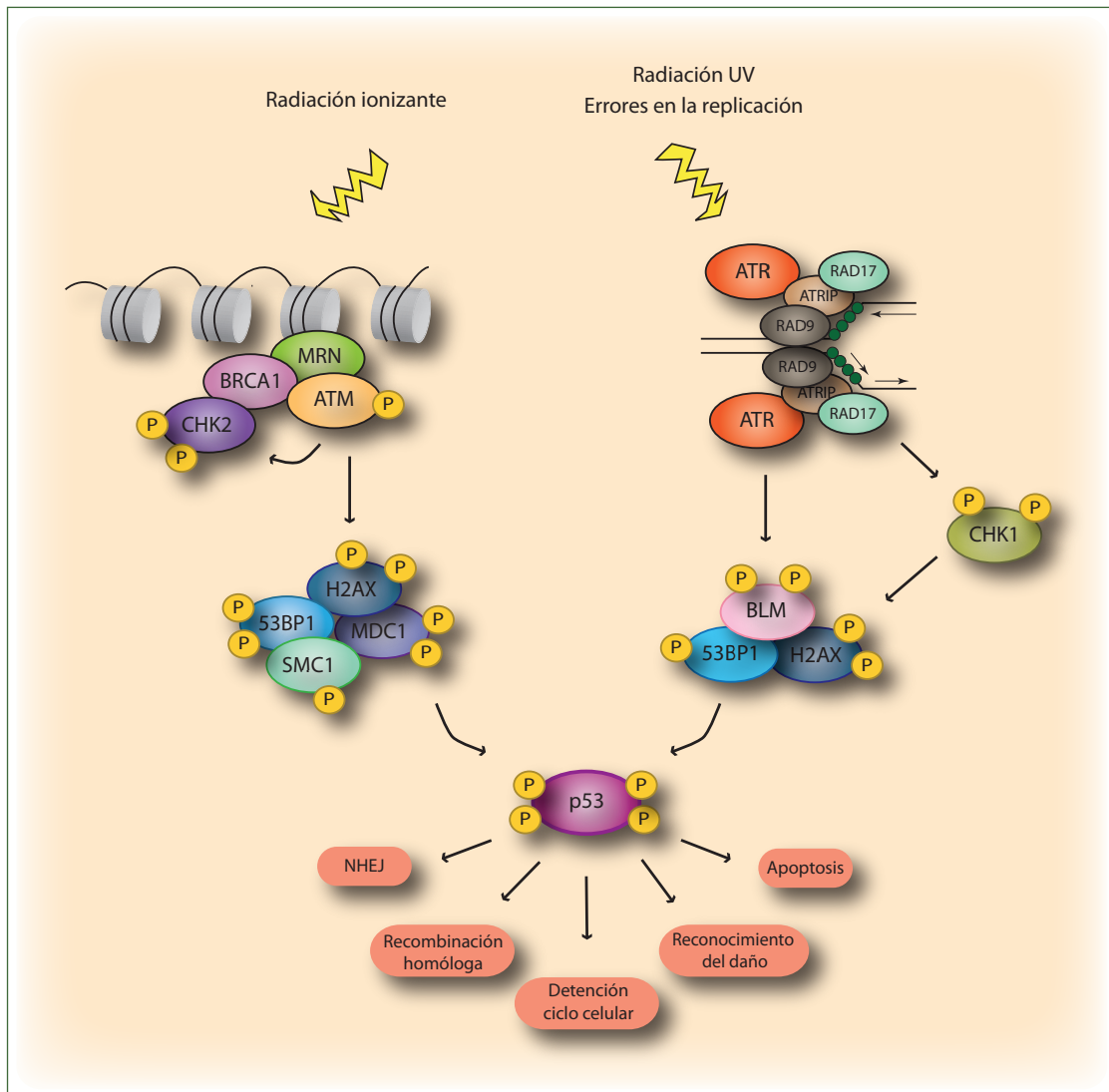
## 1.2. La activación de los mecanismos de señalización celular ocurre en respuesta al daño en el DNA

Cuando ocurre una rotura en la doble cadena de DNA, el nucleosoma, la unidad funcional básica del genoma en eucariotas, impide el acceso adecuado de las enzimas de reparación al sitio del daño. El nucleosoma está constituido por DNA y diversas proteínas, pertenecientes a la misma familia, denominadas histonas. Para superar esta barrera, las células poseen un grupo de enzimas encargadas de modificar las histonas, así como complejos proteicos que tienen la función de remodelar la cromatina. Los primeros llevan a cabo una modificación química de las histonas a través de fosforilaciones, acetilaciones o sumoirilaciones, entre otras, mientras que los complejos remodeladores de cromatina rompen los enlaces establecidos entre el DNA y las proteínas (van Attikum y Gasser, 2009). Una vez el DNA se encuentra accesible, se activan los mecanismos de transducción de la señal en respuesta al daño en el DNA, los cuales están regulados por distintas proteínas quinasas de la familia de las PIKKs (*Phosphatidylinositol-3 kinase-like kinases*), como ATM y ATR (Figura 1.2).

Estas quinasas son reclutadas al sitio de la lesión a través de la interacción directa con el DNA dañado o con determinadas proteínas, y activan a continuación a otras moléculas mediante fosforilación (Lovejoy y Cortez, 2009). ATM se activa ante la presencia de roturas en la doble cadena de DNA (Jackson, 2002; Shiloh, 2001), mientras que ATR responde ante la detención de la horquilla de replicación, uniéndose a DNA de cadena sencilla y a las proteínas ATRIP y RPA (Costanzo et al., 2003; Zou y Elledge, 2003). A partir de este momento, se activan complejas rutas de señalización celular, las cuales variarán dependiendo del tipo de lesión o el estado del ciclo celular en el que se encuentren (Stucki y Jackson, 2006). Estos mecanismos de control tienen como finalidad la detención del ciclo celular, impidiendo que se inicien nuevos orígenes de replicación, así como la inhibición de la entrada de la célula en mitosis hasta que la rotura haya sido reparada y la duplicación del genoma se haya completado. Si el proceso de reparación no finaliza con éxito, la célula entra en apoptosis (Cimprich y Cortez, 2008; Lee y Paull, 2007; Shechter et al., 2004).

## 1.3. La reparación de las roturas en la doble cadena de DNA es llevada a cabo mediante dos mecanismos diferentes

Existen dos mecanismos principales para reparar las roturas en la doble cadena de DNA en las células eucariotas con el objetivo de mantener la integridad del genoma, conocidas como recombinación homóloga y unión de extremos no homólogos. Cada uno de ellos actúa en diferentes puntos del ciclo celular, aunque se ha observado que en determinados momentos del mismo ambos procesos se encuentran activos (Burma et al., 2006; Lobrich y Jeggo, 2007).



**Figura 1.2. Cascadas de señalización celular activas tras la rotura en la doble cadena de DNA.** Cuando se produce una lesión en el DNA, se activan múltiples rutas de señalización que tienen como objetivo el mantenimiento de la integridad genómica. La señal se transmite a través de fosforilaciones (señaladas con “P” en la figura). En este proceso participan un conjunto de proteínas, entre ellas el complejo MRN (Mre11, Rad51, Nbs1), las quinasas ATM y ATR, la proteína p53, entre otras, cuya activación conduce a distintas respuestas celulares (Modificado de Sengupta y Harris, 2005).

### 1.3.1. El sistema de reparación basado en recombinación homóloga

Durante los últimos años se ha avanzado de manera significativa en el conocimiento de los mecanismos moleculares que dirigen la recombinación homóloga gracias a los estudios llevados a cabo en distintos campos como la genética, biología molecular, celular y estructural, realizados por diversos grupos de investigación, sobre todo en levaduras (Branzei y Foiani, 2008; Helleday et al., 2007).

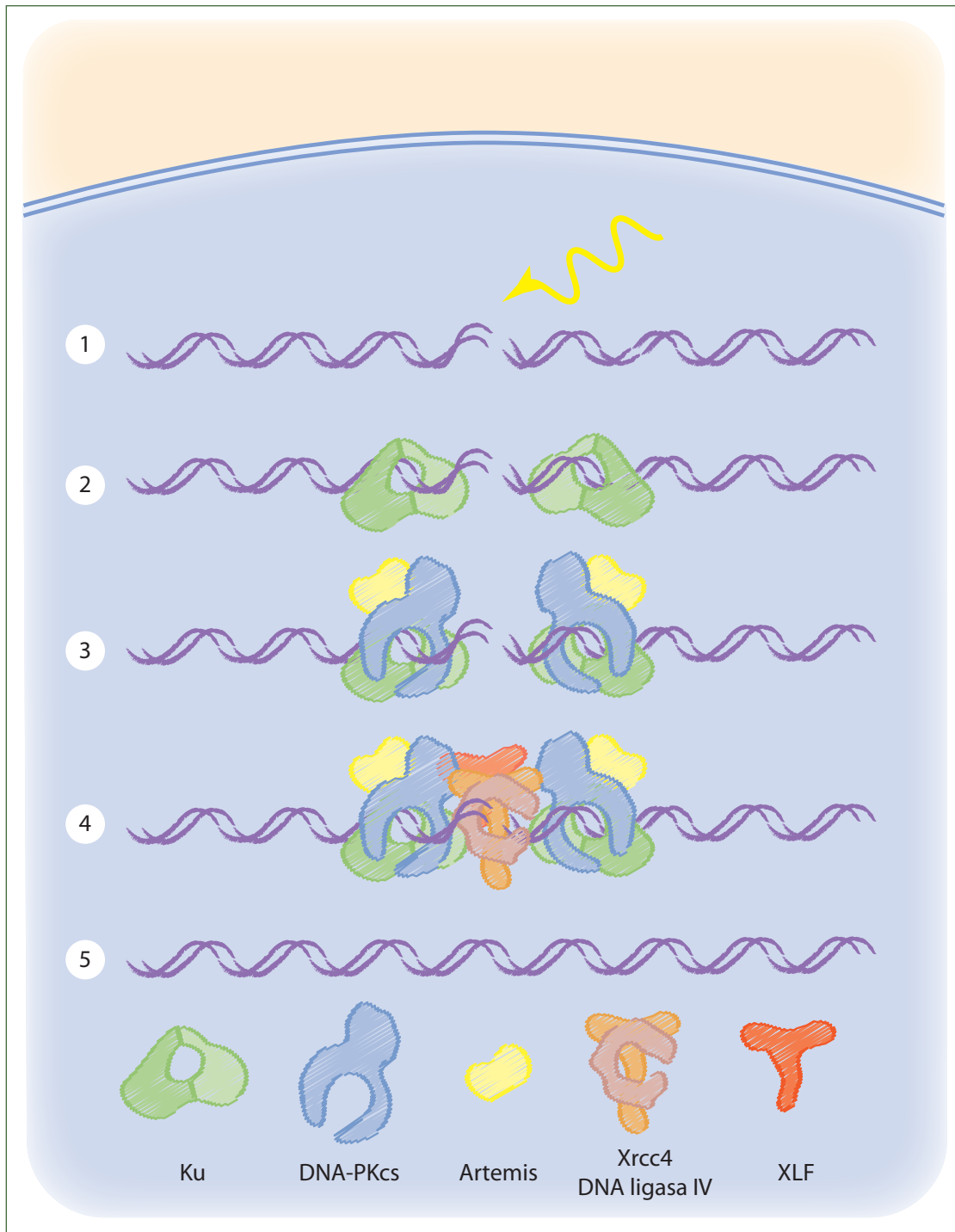


Los modelos más aceptados hasta la fecha proponen que el proceso de reparación mediante recombinación homóloga se encuentra íntimamente relacionado con la replicación del DNA. Los daños producidos en el DNA durante este proceso suponen un obstáculo para la progresión de la horquilla de replicación, con el consiguiente riesgo de que la información genética no sea transmitida fielmente a las células hijas (Segurado y Tercero, 2009). Por tanto, la reparación de las lesiones producidas en el DNA durante su replicación son críticas para el mantenimiento de la estabilidad genómica (Branzei y Foiani, 2008).

Los estudios llevados a cabo hasta el momento indican que cuando se detecta una rotura en el DNA durante la replicación, se produce una eliminación de nucleótidos a partir del extremo 5'-P del DNA. Se genera así un fragmento de DNA de cadena sencilla con el extremo 3'-OH accesible, al cual se unen diferentes proteínas constituyendo un filamento de nucleoproteína. Este filamento invade la cromátida hermana no dañada con el objetivo de encontrar una región homóloga a lo largo de la cadena. Cuando lo hace, desplaza a una de las hebras del DNA, formándose un heterodúplex entre la hebra de la molécula de DNA dañado y la cadena complementaria de su cromátida hermana. A continuación, comienza la síntesis de la nueva cadena a partir de la rotura y la hebra recién sintetizada es liberada, la cual aparea de nuevo con la región correspondiente en la cromátida donde ocurrió la lesión. Finalmente, los extremos son ligados (Cromie et al., 2001; Helleday et al., 2007). Se ha determinado que este mecanismo de reparación de DNA permanece activo durante las fases S y G<sub>2</sub> del ciclo celular exclusivamente, puesto que necesita que la cromátida hermana de aquélla que sufrió el daño actúe como molde para reparar la cadena dañada. (Branzei y Foiani, 2008; Lobrich y Jeggo, 2007). Se trata por tanto de una vía de reparación que restaura por completo la secuencia de la cadena original.

### 1.3.2. El mecanismo de reparación mediante unión de extremos no homólogos (*Non-Homologous End-Joining* o NHEJ)

La unión de extremos no homólogos es el mecanismo de reparación predominante en las células de mamífero y se encuentra activo durante todo el ciclo celular (Figura 1.3) (Burma et al., 2006; Rothkamm et al., 2003). NHEJ es un mecanismo cuyo objetivo es la restauración de la integridad estructural cromosómica, aunque ello implique la pérdida de nucleótidos. Para ello, aproxima los extremos rotos y, tras su procesamiento en el caso de que sea necesario, lleva a cabo la ligación de los mismos. Por tanto, la secuencia de la cadena original no es restaurada en este caso, ya que este proceso no utiliza una hebra complementaria que actúe como molde como en el caso de la recombinación homóloga, y por tanto se pueden producir errores en la reparación (Mahaney et al., 2009). Hasta hace no demasiado tiempo, se pensaba que el NHEJ era un proceso exclusivo de las células eucariotas, pero algunas bacterias cuyo ciclo celular transcurre mayoritariamente en fase estacionaria (y por tanto no disponen de la hebra de DNA



**Figura 1.3. Esquema del proceso de reparación mediante NHEJ.** Cuando se produce una rotura en la doble cadena de DNA, el proceso de reparación mediante la unión de extremos no homólogos es llevado a cabo en distintas etapas por un amplio conjunto de proteínas que se unen a él en distintos momentos después de producirse y detectarse la rotura. Entre ellas se encuentran la quinasa DNA-PKcs, las polimerasas  $\mu$  y  $\lambda$ , la nucleasa Artemis, XLF y el complejo Xrcc4-DNA ligasa IV.

complementaria) también presentan una ruta de reparación similar (Aravind y Koonin, 2001; Doherty et al., 2001; Weller y Doherty, 2001).

Una intensa labor investigadora llevada a cabo durante los últimos años por diversos grupos ha permitido descifrar los aspectos principales de este mecanismo de reparación. Aunque muchos detalles se desconocen o son controvertidos, existe un cierto consenso sobre su modelo básico de funcionamiento (Figura 1.3). Se ha determinado que el complejo de proteínas encargado de detectar el daño producido en el DNA es el heterodímero Ku (constituido por las subunidades Ku70 y Ku80), el cual se une a cada uno de los extremos generados con una alta afinidad y sin aparente especificidad de secuencia una vez producida una rotura en la doble cadena de DNA (Figura 1.3, 2) (Downs y Jackson, 2004). A continuación, este complejo se desplaza hacia el interior de la molécula de DNA y recluta al resto de proteínas necesarias para llevar a cabo la reparación de la lesión. Inicialmente se describió un orden de reclutamiento determinado, pero la mayoría de los modelos actuales proponen que dicho orden puede ser muy flexible (Lieber, 2008).

Se ha descrito que una de las proteínas reclutadas por el heterodímero Ku es DNA-PKcs (del inglés *DNA-dependent protein kinase catalytic subunit*), una serín/treonín quinasa de la familia de las PIKKs (Figura 1.3, 3) (Abraham, 2004; Gottlieb y Jackson, 1993). DNA-PKcs forma con Ku una holoenzima denominada DNA-PK, encargada de aproximar los extremos del DNA, al tiempo que los protege del ataque de nucleasas y de una ligación prematura (DeFazio et al., 2002; Weterings y Chen, 2008). Además, se ha descrito que la actividad quinasa de DNA-PKcs es requerida para que el proceso de NHEJ sea eficiente, aunque su función concreta no se conoce aún de manera precisa (Kurimasa et al., 1999). Se ha observado que esta quinasa fosforila a la mayoría de las proteínas implicadas en el mecanismo de reparación *in vitro*, pero se desconoce si estas fosforilaciones son necesarias *in vivo* (Yu et al., 2008b; Yu et al., 2003). También se ha descrito que la autofosforilación *in vitro* de DNA-PKcs tiene como consecuencia la pérdida de la actividad quinasa y la disociación del complejo del DNA, por lo que se cree que las fosforilaciones pueden ser necesarias para el desensamblaje de las proteínas del DNA una vez reparada la lesión (Chan et al., 2002).

Se ha demostrado que las especies reactivas de oxígeno y la radiación ionizante son capaces de producir un conjunto heterogéneo de roturas en la doble cadena de DNA. Según el tipo de extremos que hayan sido generados tras la rotura, puede ser necesaria una fase previa a la ligación en la que tiene lugar un procesamiento de los extremos, de manera que las proteínas implicadas pueden ser diferentes en cada caso concreto. Algunas de estas proteínas son Artemis, una nucleasa que se une a DNA-PKcs con el fin de escindir nucleótidos de los extremos del DNA (Figura 1.3, 3), o las polimerasas  $\mu$  y  $\lambda$ , las cuales se encargan de sintetizar nuevos nucleótidos. Finalmente, la ligación de los extremos es llevada a cabo por un complejo constituido por las proteínas Xrcc4 (del

inglés *X-ray cross-complementing protein 4*) y DNA ligasa IV, y dependiendo del tipo de extremos a reparar también pueden intervenir otros factores como XLF (del inglés *Xrcc4-like factor*), también conocido como Cernunnos (Figura 1.3, 4 y 5) (Mahaney et al., 2009).

El estudio de la maquinaria de reparación que participa en NHEJ constituye una diana importante para tratamientos dirigidos a células tumorales sensibles a la radiación, debido a su papel crucial en la reparación de roturas de doble cadena en el DNA (Lee et al., 2003). Hasta el momento, el desarrollo de inhibidores frente a DNA-PKcs ha sido la estrategia esencial a seguir, pero el complejo de ligación puede ser una diana alternativa. Esto es debido a que son proteínas específicas de las rutas NHEJ y recombinación V(D)J, a diferencia de DNA-PKcs y Ku que también se encuentran implicadas en funciones celulares adicionales tales como el mantenimiento de los telómeros, la transcripción y la apoptosis (Downs y Jackson, 2004).

### **1.4. El proceso de ligación durante la reparación mediante NHEJ es llevado a cabo por el complejo Xrcc4-DNA ligasa IV**

La ligación es el proceso central del mecanismo de reparación del DNA mediada por NHEJ y requiere la acción conjunta de dos proteínas denominadas en mamíferos Xrcc4 y DNA ligasa IV (Wu et al., 2009). Este complejo está presente en todos los eucariotas, y se expresa constitutivamente en la célula (Grawunder et al., 1998b). En la levadura *S. cerevisiae* estas proteínas se denominan Lif1p (del inglés *Ligase interacting factor 1*) y Lig4p (del inglés *Ligase 4*), respectivamente. El complejo participa exclusivamente en la ruta de reparación de DNA conocida como NHEJ, así como en el proceso de recombinación V(D)J en las células especializadas del sistema inmune (Critchlow et al., 1997; Grawunder et al., 1998a; Li et al., 1995). Xrcc4 es una proteína que no presenta actividad enzimática, y entre las funciones que se le atribuyen se encuentran la estabilización de DNA ligasa IV, a la vez que estimula su adenilación previa al proceso de ligación y la protege de la degradación (Bryans et al., 1999; Modesti et al., 1999).

Recientemente, ha sido descubierto un nuevo factor que participa también en NHEJ y al que se ha denominado XLF y Cernunnos (Ahnesorg et al., 2006; Buck et al., 2006). La función más importante asignada hasta el momento a XLF es la estimulación *in vitro* de la actividad ligasa del complejo Xrcc4-DNA ligasa IV en la unión de extremos desapareados y no cohesivos, sugiriendo que XLF puede regular su actividad en determinadas situaciones (Gu et al., 2007a; Gu et al., 2007b; Hentges et al., 2006; Li et al., 2008; Lu et al., 2007; Tsai et al., 2007). Se ha observado que la ligasa es necesaria para la formación y función de este complejo ternario, pero existe una gran controversia sobre si ésta dirige a XLF al lugar de la lesión en el DNA. Parece que XLF interacciona con ella aunque débilmente a través del dominio C-terminal de la ligasa (Andres et al.,

2007). También se ha descrito que XLF y Xrcc4 interaccionan *in vitro* a través de sus correspondientes dominios N-terminal, aunque esta última no es requerida para el reclutamiento de XLF a los lugares de daño en el DNA *in vivo* (Ahnesorg et al., 2006; Li et al., 2008; Yano et al., 2008).

Xrcc4, XLF y DNA ligasa IV interaccionan además con otros factores del mecanismo de reparación NHEJ, así como con DNA (Costantini et al., 2007; Mari et al., 2006; Modesti et al., 1999; Nick McElhinny et al., 2000). El complejo Xrcc4-DNA ligasa IV parece ser reclutado por Ku al sitio donde se localiza la lesión en el DNA, y se cree que XLF puede hacerlo posteriormente, también mediante la unión al complejo Ku-DNA. XLF es a continuación estabilizado y retenido en el punto de rotura gracias a la interacción con el complejo Xrcc4-DNA ligasa IV (Yano et al., 2008), aunque también se ha propuesto que DNA-PKcs favorece la retención del complejo en los extremos del DNA (Costantini et al., 2007).

Se ha determinado que Xrcc4, DNA ligasa IV y XLF se encuentran reguladas mediante fosforilación por DNA-PKcs, al igual que el resto de las proteínas de este mecanismo de reparación. Xrcc4 y XLF son fosforiladas *in vitro* e *in vivo* en sus respectivos dominios C-terminal, aunque parece que la fosforilación no es esencial para un correcto desarrollo del NHEJ y de la recombinación V(D)J, por lo que el significado biológico que puedan tener estos procesos permanecen aún sin esclarecer, aunque parecen producir un incremento en la eficiencia de la ligación. (Critchlow et al., 1997; Drouet et al., 2005; Lee et al., 2004; Matsumoto et al., 2000; Yu et al., 2008b). Además, se cree que la disociación de las proteínas del DNA se puede producir por dichos fenómenos de fosforilación una vez finalizada la reparación.

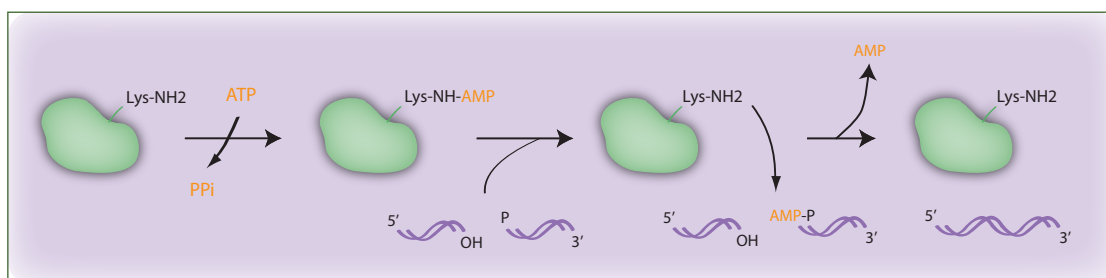
Se ha observado que las células que carecen o tienen mutaciones en alguna de las proteínas Xrcc4, DNA ligasa IV y XLF, muestran defectos en el proceso de recombinación V(D)J que desemboca en inmunodeficiencias y retrasos en el desarrollo (O'Driscoll et al., 2001). Además, presentan una extremada sensibilidad a la radiación ionizante e inestabilidad genómica, lo cual incrementa la susceptibilidad a desarrollar un proceso cancerígeno (O'Driscoll et al., 2001; Riballo et al., 1999; Smith et al., 2003).

### **1.5. Estructura de la DNA ligasa IV, la ligasa específica del proceso de reparación mediante unión de extremos no homólogos**

La enzima DNA ligasa IV es un polipéptido de 108 kDa de peso molecular que pertenece a la familia de DNA ligasas dependientes de ATP. Las DNA ligasas son nucleotidiltransferasas (NTasas), una familia de fosfotransferasas con motivos conservados en la secuencia que incluye tanto DNA ligasas como RNA ligasas. Estas enzimas utilizan un cofactor, NAD<sup>+</sup> o ATP, para catalizar la formación de un enlace fosfodiéster entre dos oligonucleótidos (Ellenberger y Tomkinson, 2008). Las DNA ligasas dependientes

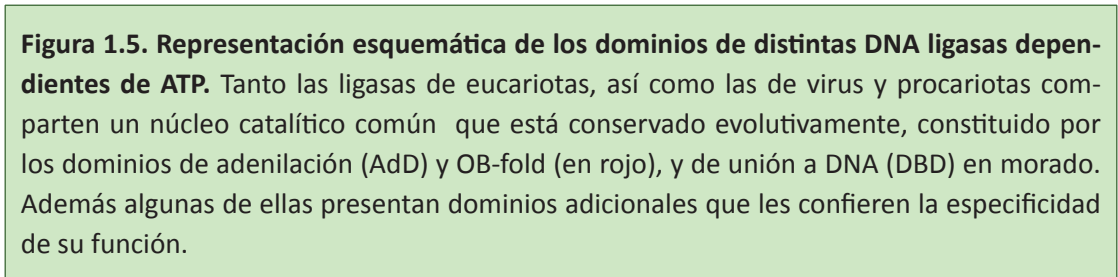
de NAD<sup>+</sup> se localizan en todos los organismos excepto en eucariotas, mientras que las ATP ligasas se encuentran tanto en eucariotas como en procariotas y virus (Shuman, 2009). El proceso de ligación se encuentra dividido en tres fases (Figura 1.4). En primer lugar, un grupo adenilato (AMP) es transferido desde una molécula de ATP (en el caso de las ATP ligasas) o NAD<sup>+</sup> (en las NAD ligasas) a un residuo lisina localizado en el sitio activo de la enzima. A continuación, el AMP es transferido desde la ligasa al grupo 5'-P de uno de los extremos del DNA. La adenilación del DNA tiene como consecuencia la activación del grupo 5'-P, produciéndose un ataque nucleofílico por parte del grupo 3'-OH presente en el otro extremo del DNA que desplaza el AMP, uniéndose así covalentemente los extremos de las dos hebras de DNA. Esta reacción es energéticamente favorable y esto la hace completamente irreversible (Ellenberger y Tomkinson, 2008).

En eucariotas se han descrito tres familias de DNA ligasas dependientes de ATP: DNA ligasa I, III y IV, relacionadas en estructura y secuencia (Figura 1.5). Todas comparten una región catalítica que está muy conservada incluso en virus y procariotas, y que comprende tres dominios que interaccionan con DNA: un dominio N-terminal llamado de unión a DNA (DBD, del inglés **DNA binding domain**) esencial para finalizar con éxito la ligación, el dominio nucleotidiltransferasa o de adenilación (AdD, del inglés **Adenylation domain**) que une el ATP, y un dominio de unión a oligonucleótidos y oligosacáridos denominado dominio (OB)-fold (del inglés **Oligomer Binding-fold**) que participa en la transferencia del AMP al residuo lisina localizado en el sitio activo del dominio de adenilación (Figura 1.5). Los dominios adyacentes al centro catalítico de las DNA ligasas eucariotas son diferentes en cada una de las familias, y son los que les confieren la especificidad para llevar a cabo determinadas funciones. Estas regiones suelen estar implicadas en interacciones específicas con otras proteínas de reparación y/o replicación (Ellenberger y Tomkinson, 2008). En la proteína DNA ligasa IV, la región C-terminal adyacente al centro catalítico está constituida por dos dominios BRCT (del inglés **BR**east **C**ancer **A**sociated **g**ene **1 C-terminal**). Estos dominios están conservados en multitud de proteínas relacionadas con señalización celular y reparación de DNA



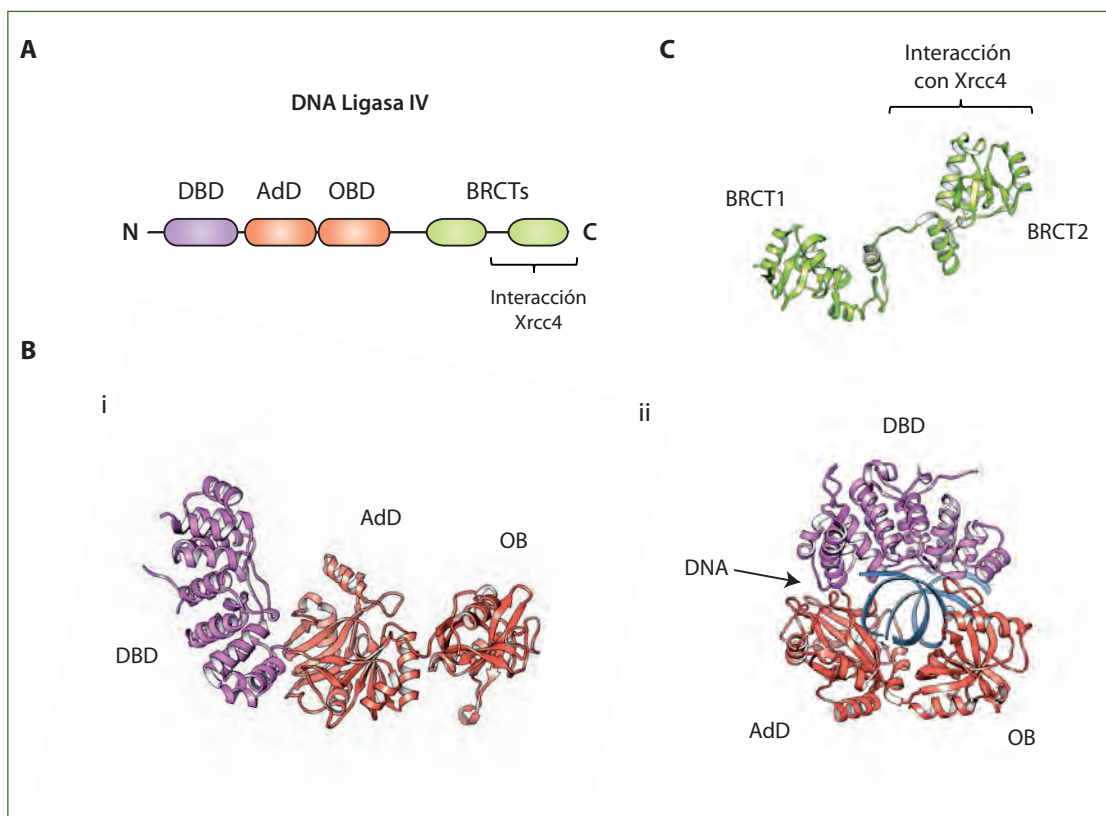
**Figura 1.4. Mecanismo de ligación en las ligasas dependientes de ATP.** La reacción se divide en tres etapas diferenciadas. El AMP es transferido desde una molécula de ATP al sitio activo de la enzima, y después a un extremo 5'-P de la molécula de DNA. Tras un ataque nucleofílico del extremo 3'-OH sobre el extremo 5'-P se produce la unión covalente entre los dos extremos de DNA.





La biología estructural de las DNA ligasas ha avanzado de forma significativa en los últimos años con la resolución de la estructura atómica de los dominios catalíticos de multitud de ligasas (Figura 1.6). Aunque la estructura cristalográfica de la proteína DNA ligasa IV no ha sido aún determinada, las estructuras resueltas de otras ligasas han proporcionado nuevos datos sobre la organización de los dominios catalíticos y su regulación. Las condiciones en las que se han resuelto estas estructuras varían desde aquéllas no unidas a ligando (Ho et al., 1997; Pascal et al., 2006), en presencia de ATP (Odell et al., 2000; Pascal et al., 2004) o unidas a DNA (Odell y Shuman, 1999) (Pascal et al., 2004). En estas estructuras se puede observar que todas ellas, a excepción de las ligasas virales más simples, forman una estructura abierta en ausencia de DNA que adopta forma de C cuando se encuentran formando complejo con el DNA mellado, rodeándolo completamente (Figura 1.6, B). Esto es debido a su arquitectura con tres dominios que constituyen el núcleo catalítico, y que le confiere la suficiente flexibilidad para abrirse y cerrarse alrededor del DNA (Ellenberger y Tomkinson, 2008).





**Figura 1.6. Esquema de los dominios de DNA ligasa IV y estructuras atómicas resueltas.**

**A)** Esquema de dominios de la proteína. **B)** Estructuras atómicas de los dominios catalíticos resueltos en otras DNA ligasas dependientes de ATP, homólogas a DNA ligasa IV, en ausencia o presencia de DNA: i, DNA ligasa de *Sulfolobus solfataricus* (Pascal et al., 2006) (archivo PDB 2HIV); ii, estructura de la proteína DNA ligasa I humana (Pascal et al., 2004) (archivo PDB 1X9N). En esta última se aprecia que el núcleo catalítico conservado (dominios AdD y OB, en rojo) contacta con los extremos de DNA dañado durante el proceso de ligación. Un tercer dominio (DBD, en morado) se encarga de formar el anillo que rodea completamente al DNA. **C)** Estructura atómica de los dominios BRCT de la proteína DNA ligasa IV humana (Wu et al., 2009) (parte del PDB 3II6, el cual se muestra completo en la Figura 1.7).

A pesar de todas las estructuras resueltas, la única región conocida en la DNA ligasa IV son los dominios BRCT junto con el fragmento que las conecta y a través del cuál interacciona con la proteína Xrcc4 (Dore et al., 2006; Sibanda et al., 2001; Wu et al., 2009). Los dominios BRCT se han resuelto como parte de un complejo con varios dominios de Xrcc4 (Figura 1.6, C y Figura 1.8). En esta estructura se aprecia que los dominios BRCT presentan las mismas características en el plegamiento que otros dominios de la familia previamente resueltos, ya que están conservados, y están constituidas por cuatro láminas  $\beta$  paralelas en el centro flanqueadas por tres alfa-hélices (Dore et al., 2006; Wu et al., 2009).

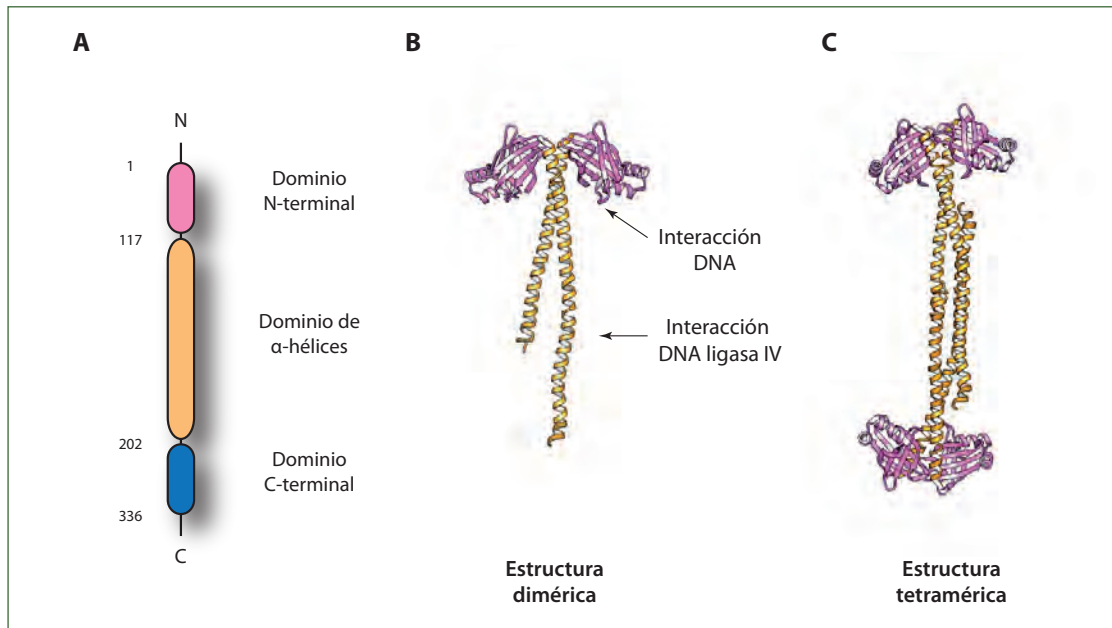
Como se ha comentado anteriormente, hasta la fecha no se ha resuelto ninguna estructura que contenga los dominios catalíticos de DNA ligasa IV, por lo que se desconoce actualmente la disposición de esta región relativa a los dominios BRCT y a Xrcc4.

## 1.6. Características estructurales de Xrcc4

Xrcc4 es una proteína altamente conservada en todos los eucariotas, y no muestra homología con ninguna otra proteína descrita previamente. Xrcc4 fue el tercer gen identificado cuyo producto participa en la reparación mediante NHEJ y en la recombinación V(D)J, tras el descubrimiento de DNA-PKcs y Ku (Li et al., 1995). Como se ha comentado anteriormente, su principal función parece ser la estabilización de la proteína DNA ligasa IV, puesto que forman un complejo constitutivo en la célula. Además, estimula su adenilación (Modesti et al., 1999), promoviendo interacciones estables con DNA, así como protegiéndola de posibles fenómenos de degradación (Bryans et al., 1999; Foster et al., 2006). Durante el desarrollo de esta tesis se han estudiado la proteína de ratón por un lado, y la de origen humano en complejo con la proteína DNA ligasa IV por otro. La proteína humana contiene 336 residuos y tiene un peso molecular de 38 kDa, mientras que la de ratón, homóloga en un 76% a la primera, contiene diez residuos menos. Se ha descrito que Xrcc4 existe en la célula como un homodímero funcional, y distintos grupos han resuelto diversas estructuras de la proteína mediante cristalografía de rayos X (Figura 1.7). En todos los casos, se trata de moléculas truncadas que comprenden alrededor de las dos terceras partes de la proteína. Estas estructuras muestran que Xrcc4 está constituida por un dominio N-terminal globular de topología  $\beta$  sándwich y un dominio de alfa-hélices superenrolladas paralelas de gran longitud (alrededor de 120 Å) en la parte central (*coiled-coil*). Además, la proteína dispone de un dominio C-terminal que comprende alrededor de un tercio de la molécula y cuya estructura y función permanecen aún desconocidas (Junop et al., 2000) (fichero PDB 1FU1, Figura 1.7, A y B). También se ha determinado mediante estudios bioquímicos y estructurales que dos dímeros pueden interaccionar a través del dominio de *coiled-coil* para dar lugar a un tetrámero, si bien los estudios realizados sugieren que estos tetrámeros no tienen relevancia funcional *in vivo* (Figura 1.7, C) (Junop et al., 2000; Leber et al., 1998).

Algunas de las estructuras atómicas de Xrcc4 se han resuelto en presencia de algún fragmento de un dominio BRCT de la DNA ligasa IV (fichero PDB 1IK9, Figura 1.8, A) o más recientemente junto con ambos dominios (fichero PDB 3II6, Figura 1.8, B) (Sibanda et al., 2001; Wu et al., 2009). También ha sido determinada la estructura del complejo ortólogo en *S. cerevisiae* (Lif1p-Lig4p) (fichero PDB 1Z56, Figura 1.8, C) (Dore et al., 2006). En todas estas estructuras se puede observar que la región de unión de los dominios BRCT de la DNA ligasa IV interacciona con la forma dimérica de Xrcc4 a través del dominio *coiled-coil* (Junop et al., 2000; Sibanda et al., 2001; Wu et al., 2009). La estequiometría de la reacción es 2:1 en todos los casos, por lo que se cree que la interacción de Xrcc4 con la ligasa impide la formación de tetrámeros, hecho que ha sido confirmado mediante experimentos llevados a cabo *in vivo* (Modesti et al., 2003).

Se ha observado que Xrcc4 puede interaccionar con moléculas de DNA de gran lon-



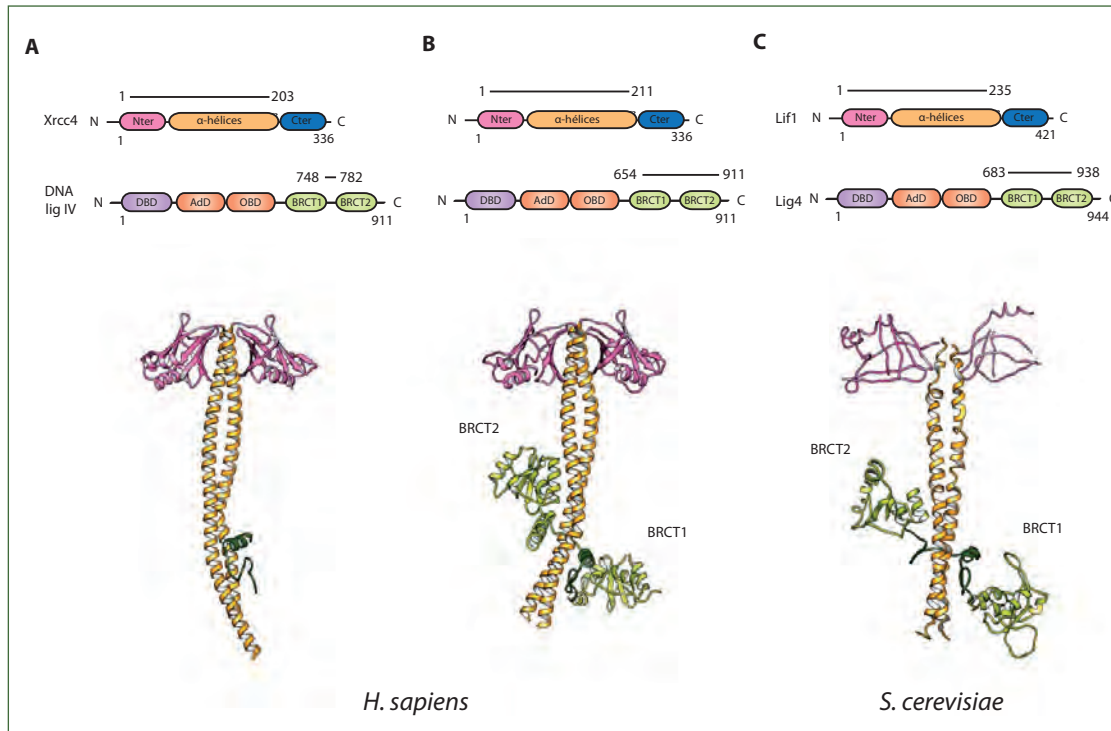
**Figura 1.7. Características estructurales de la proteína Xrcc4.** **A)** Representación esquemática de los dominios de Xrcc4. **B)** Estructura atómica del dímero de Xrcc4 de origen humano resuelto mediante cristalografía de rayos X, el cual carece del dominio C-terminal (Junop et al., 2000) (fichero PDB 1FU1). **C)** Estructura tetramérica de Xrcc4, caracterizada tanto bioquímicamente como mediante técnicas estructurales (Junop et al., 2000).

gitud (mayores de 100 pb), tanto en presencia como en ausencia de la ligasa, aunque se desconoce la región de interacción (Chen et al., 2000; Modesti et al., 1999). Se ha propuesto que puede hacerlo a través de la región del dominio de *coiled-coil* más próxima a la base del dominio N-terminal, ya que es la única que muestra un potencial electrostático positivo (Junop et al., 2000; Modesti et al., 1999). Otros autores sugieren que la interacción podría localizarse en la región *coiled-coil* (Sibanda et al., 2001). Por último, otros estudios han determinado que Xrcc4 une DNA de forma cooperativa, por lo que es posible que posea diferentes puntos de unión a lo largo de la molécula (Junop et al., 2000; Modesti et al., 1999).

A pesar de todas estas estructuras resueltas, aún no se conoce la estructura de la proteína Xrcc4 completa, así como la del complejo que forma con la DNA ligasa IV, algo esencial para entender el proceso de ligación.

### 1.7. XLF/Cernunnos interviene en determinados procesos de reparación mediante NHEJ

XLF o Cernunnos es una proteína de reciente descubrimiento que participa en el mecanismo de reparación NHEJ y en la recombinación V(D)J (Ahnesorg et al., 2006; Buck et al., 2006; Sekiguchi y Ferguson, 2006). XLF fue identificado en el año 2006 simultáneamente por dos grupos de investigación distintos utilizando abordajes com-



**Figura 1.8. Estructuras atómicas de regiones de Xrcc4 interaccionando con DNA ligasa IV.**

Xrcc4 interacciona con los dominios BRCT de DNA ligasa IV a través de su dominio *coiled-coil*. **A)** y **B)** Estructuras resueltas en humanos (archivos PDB 1IK9 y 3II6, respectivamente), las cuales contienen parte o la totalidad del dominio BRCT de la DNA ligasa IV (Junop et al., 2000; Wu et al., 2009). La proteína Xrcc4 tampoco se ha resuelto completa, careciendo de su dominio más C-terminal. **C)** Estructura de un ortólogo del complejo humano resuelto en la levadura *S. cerevisiae* (Dore et al., 2006) (archivo PDB 1Z56).

pletamente diferentes. Por un lado, Buck y colaboradores descubrieron esta proteína estudiando pacientes que manifestaban inmunodeficiencias combinadas con sensibilidad a la radiación, síntomas que estaban relacionados con defectos en NHEJ (Buck et al., 2006). Mediante estudios de complementación, aislaron el gen que estaba mutado en estos pacientes y lo denominaron Cernunnos. Por otro lado, Ahnesorg y colaboradores aislaron este nuevo factor mediante ensayos de doble híbrido cuando intentaban descubrir proteínas que interaccionaban con el complejo Xrcc4-DNA ligasa IV (Ahnesorg et al., 2006). Determinaron que este factor interaccionaba con Xrcc4 y encontraron que este gen se hallaba inactivo en pacientes con inmunodeficiencias combinadas severas, las cuales no eran debidas a mutaciones en ningún otro factor implicado en NHEJ conocido hasta ese momento. Las predicciones de estructura sugerían que la proteína descubierta podría tener una estructura similar a Xrcc4 a pesar de su baja homología en la secuencia, y por ello se denominó XLF (*Xrcc4-like-factor*).

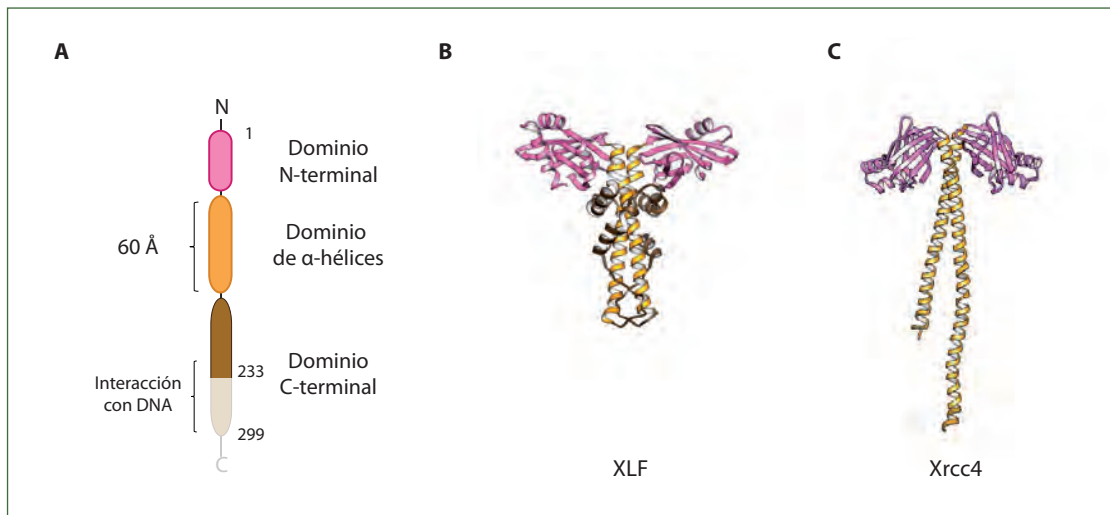
Desde su identificación, se ha realizado un importante avance en el conocimiento de la función de XLF en el proceso de NHEJ, determinándose que es esencial en algunos procesos de reparación mediante NHEJ cuando los extremos del DNA no pueden

ser alineados correctamente porque son incompatibles. Se ha determinado que XLF está conservado funcional y evolutivamente en eucariotas en los 220 primeros aminoácidos, especialmente en vertebrados, pero sólo en pequeños motivos de secuencia (Ahnesorg et al., 2006; Callebaut et al., 2006; Hentges et al., 2006).

La proteína XLF de origen humano tiene un peso molecular de 33 kDa y está constituida por tres dominios (Figura 1.9, A). Recientemente, dos estudios independientes resolvieron dos estructuras atómicas parciales de XLF mediante cristalografía de rayos X, carentes ambas del dominio C-terminal (ficheros PDB 2QM4 y 2R9A) (Andres et al., 2007; Li et al., 2008). Estos estudios muestran que la estructura terciaria de XLF constituye un homodímero estructuralmente similar a Xrcc4, a pesar de su escasa homología en la secuencia primaria de aminoácidos, principalmente en la región N-terminal y parte del dominio *coiled-coil* (Figura 1.9, B) (Ahnesorg et al., 2006; Andres et al., 2007; Callebaut et al., 2006; Li et al., 2008). El dominio *coiled-coil* en XLF es de tan sólo 60 Å de longitud, y al igual que en Xrcc4, se trata de la región de interacción entre los dos monómeros. Sin embargo, estas hélices se repliegan sobre sí mismas en una región formada por alfa-hélices y regiones desordenadas que se aproximan de nuevo a la región N-terminal. Por tanto, el dominio C-terminal de XLF, del cual no se conoce su estructura, en este caso se dispone cerca del dominio N-terminal, al contrario que en Xrcc4. Estudios de predicción de estructura parecen indicar que esta región no forma una estructura definida (Andres et al., 2007; Hentges et al., 2006; Li et al., 2008). Sin embargo, se ha determinado que los últimos 75 residuos más próximos al dominio C-terminal son esenciales para la unión de XLF a DNA. Esta interacción, al igual que en el caso de Xrcc4, es dependiente de la longitud del DNA y se ve favorecida por la interacción con Ku (Yano et al., 2008). También se ha determinado que el dominio C-terminal es necesario para la activación del complejo Xrcc4-DNA ligasa IV (Andres et al., 2007).

Se ha determinado que XLF y Xrcc4 interaccionan a través de sus dominios N-terminal (Ahnesorg et al., 2006; Li et al., 2008; Yano et al., 2008). Cuando se descubrió XLF se pensó en un principio que el monómero podría interaccionar con un monómero de Xrcc4 formando un heterodímero, y que esta disposición podría constituir algún mecanismo de regulación del proceso de ligación. Sin embargo, cuando se resolvió la estructura atómica se observó que no era compatible con esa disposición (Li et al., 2008).

A pesar de los avances producidos en los últimos años, existen muchas incógnitas por desvelar en el campo de la reparación mediante NHEJ. Ahondar en el conocimiento de las estructuras de los complejos proteicos y su interacción con el DNA ayudará a entender mejor el proceso de reparación, oscuro aún en algunos de los pasos y controvertido en muchos otros.



**Figura 1.9. Descripción estructural de XLF.** **A)** Representación esquemática de los diferentes dominios que constituyen XLF. **B)** Una de las dos estructuras atómicas de XLF resueltas mediante cristalografía de rayos X (Li et al., 2008) (fichero PDB 2R9A). **C)** Estructura atómica de Xrcc4 (Junop et al., 2000) (fichero PDB 1FU1). Xrcc4 y XLF comparten ciertas características estructurales, pero son muy diferentes entre sí.

## 1.8. La microscopía electrónica aplicada a la reconstrucción tridimensional de partículas individuales

La microscopía electrónica de partículas individuales es una técnica que permite el estudio de proteínas y complejos macromoleculares utilizando pequeñas cantidades de muestra y sin necesidad de obtener cristales. Debido a estas características, se decidió emplear esta técnica para el estudio estructural de las proteínas Xrcc4 y DNA ligasa IV implicadas en el mecanismo de reparación NHEJ.

### 1.8.1. Perspectiva de la microscopía electrónica de partículas individuales

Las secuencias de aminoácidos que constituyen las proteínas se pliegan de una forma determinada en cada una de ellas para formar dominios compactos con estructuras específicas que constituyen las regiones funcionales o sitios activos. Las propiedades funcionales de las proteínas y los complejos macromoleculares que se forman en la célula dependen de su estructura tridimensional, por lo que su determinación es esencial para comprender la función que realizan (Branden y Tooze, 1999; Gómez-Moreno Calera y Sancho Sanz, 2003; Sali et al., 2003).

Existe una gran variedad de técnicas para el análisis estructural de proteínas, como la cristalografía de rayos X (Ilari y Savino, 2008), resonancia magnética nuclear (Grzesiek y Sass, 2009), microscopía electrónica (Cheng y Walz, 2009), SAXS (*Small Angle X-ray Scattering*) (Neylon, 2008), tomografía electrónica (Jonic et al., 2008), entre otras. Cada una de ellas permite alcanzar un determinado nivel de resolución, pro-



porcionando distintos grados de información estructural del espécimen bajo estudio. Por este motivo, es muy interesante que en la actualidad se esté avanzando hacia el desarrollo de metodologías que combinen la información estructural obtenida a partir de diversas técnicas a distintos niveles de resolución (Nickell et al., 2006; Russel et al., 2009).

En este contexto, la microscopía electrónica de partículas individuales tiene como finalidad la determinación de la estructura tridimensional de una molécula o complejo macromolecular a partir de sus imágenes de proyección obtenidas en el microscopio. Existen dos requerimientos básicos para realizar un estudio estructural mediante microscopía electrónica. Por un lado, la molécula ha de tener un tamaño mínimo para que pueda ser observada al microscopio, y por otro, la molécula ha de existir en muchas copias idénticas en la muestra bajo estudio (Cheng y Walz, 2009).

Las ventajas que presenta la microscopía electrónica sobre otras técnicas es que permite estudiar, con escasa cantidad de muestra, especímenes biológicos de un amplio rango de tamaños, desde estructuras celulares a moléculas individuales, proporcionando información desde baja a alta resolución (3,8-40Å) (Frank, 2006; Lucic et al., 2005). Además, permite obtener información estructural sin necesidad de generar cristales previamente, a diferencia de la cristalografía electrónica o la cristalografía de rayos X, por lo que se pueden estudiar complejos macromoleculares que no cristalizan debido a su flexibilidad o a su elevado tamaño. Otra gran ventaja es que se pueden estudiar las proteínas en su estado nativo y en su tampón fisiológico. Por otro lado, también permite analizar complejos que presenten heterogeneidad de composición o conformación, hecho que dificulta el análisis estructural mediante las técnicas tradicionales (Spahn y Penczek, 2009). En microscopía electrónica el problema de la variabilidad conformacional puede ser afrontado separando las distintas poblaciones computacionalmente. Por tanto, la obtención de estructuras tridimensionales mediante microscopía electrónica puede ser particularmente útil en la observación de moléculas que presentan heterogeneidad conformacional, aunque es necesario aún un mayor desarrollo de nuevas metodologías para poder lidiar con situaciones complejas de gran variabilidad estructural (Spahn y Penczek, 2009).

La microscopía electrónica presenta, sin embargo, limitaciones y desventajas frente a otras técnicas estructurales, especialmente en cuanto a la resolución alcanzable. Esta resolución está limitada por distintos factores, como diversos defectos de las lentes, el bajo contraste de las imágenes obtenidas a partir de especímenes biológicos, el efecto de la radiación sobre las muestras y las técnicas de preparación del espécimen. Todos estos factores impiden que se alcance la resolución teórica a la que puede llegar la microscopía electrónica, que es resolución atómica. Sin embargo, en los últimos años se ha producido un gran desarrollo de nuevos microscopios y dispositivos de toma de imágenes, así como un incremento en la velocidad y recursos computacionales y el

desarrollo de nuevos algoritmos y programas informáticos. Estos factores, junto con la automatización en muchas etapas del procesamiento de partículas individuales, han hecho posible que aumente drásticamente el número de estructuras resueltas mediante microscopía electrónica y también la resolución alcanzada, hasta la obtención de las primeras estructuras a resolución quasiatómica (menos de 4 Å) de varios virus icosaédricos (Jiang et al., 2008; Yu et al., 2008a; Zhang et al., 2008). Sin embargo, aunque continúa suponiendo un reto bajar de los 10 Å de resolución en moléculas asimétricas, con baja simetría o de pequeño tamaño, a día de hoy existen ya algunos ejemplos de ello. Entre ellos se encuentran el ribosoma, cuya estructura ha sido resuelta a 6,7 Å (Villa et al., 2009), GroEL a 6 Å (Ludtke et al., 2008) y muy recientemente la chaperona Mm-Cpn60 a 4,3 Å (Zhang et al., 2010). En este último caso, los autores han sido capaces además de trazar por primera vez un modelo atómico utilizando únicamente la información proveniente de la estructura obtenida mediante microscopía electrónica y la secuencia de la proteína.

A pesar de que se ha avanzado mucho en este campo en los últimos años, todavía existen muchos problemas metodológicos por resolver, como son la estandarización de los métodos utilizados en la determinación de las estructuras, la dificultad para estudiar moléculas pequeñas y asimétricas, encontrar un criterio común para validar la estructura generada y la necesidad de desarrollar nuevos métodos para resolver la heterogeneidad estructural.

### 1.8.2. Introducción a la metodología empleada en la obtención de una estructura tridimensional mediante microscopía electrónica

La determinación de la estructura tridimensional de moléculas biológicas mediante microscopía electrónica de partículas individuales comienza por la obtención de una muestra purificada y homogénea en solución. En microscopía electrónica es necesario en principio que todas las partículas observadas en el microscopio electrónico provengan de la misma molécula, y además en un solo estado conformacional. Sin embargo, se están desarrollando en la actualidad herramientas informáticas y computacionales, así como técnicas bioquímicas que permitan la separación de moléculas en distintas conformaciones y complejos en equilibrio (Kastner et al., 2008; Kelly et al., 2008a; Kelly et al., 2008b; Leschziner y Nogales, 2007).

La muestra a estudiar se deposita sobre un soporte denominado rejilla, la cual ha sido recubierta previamente con una fina capa de carbón. Las moléculas de la muestra son adsorbidas y se disponen sobre la rejilla preferiblemente en todas las orientaciones posibles, aunque esto depende de la forma de la molécula y la distribución de cargas. A continuación, las rejillas han de ser tratadas previamente a su visualización al microscopio electrónico, ya que de no ser así, las condiciones de vacío de la columna del equipo y la radiación electrónica degradarían la estructura biológica (Cheng y Walz,



2009; Llorca, 2005). Existen tres técnicas de preparación del espécimen: tinción negativa, vitrificación y criotinción (Frank, 2006; Llorca, 2005). La tinción negativa consiste en cubrir la muestra con un agente de tinción, que suele ser una sal de un metal pesado, como el acetato de uranilo o formiato de uranilo. Su función es proteger a la muestra de la radiación y el vacío de la columna del microscopio, a la vez que aumenta el contraste de las imágenes obtenidas. La resolución de esta técnica está limitada en torno a los 20 Å por el tamaño del grano del agente de tinción, el cual no puede penetrar en determinadas zonas de la molécula (Cheng y Walz, 2009; Llorca, 2005). La criomicroscopía electrónica es una técnica que consiste en congelar la muestra a gran velocidad en etano líquido, por lo que el hielo que se forma es vítreo y no daña al espécimen (Dubochet et al., 1988; van Heel et al., 2000). La muestra se mantiene y se visualiza congelada a temperatura de nitrógeno líquido (alrededor de -180 °C), lo cual asegura su preservación e hidratación en unas condiciones cercanas a las fisiológicas durante la exposición a los electrones (Cheng y Walz, 2009; van Heel et al., 2000). Las imágenes obtenidas mediante criomicroscopía proporcionan información estructural de toda la molécula, además de permitir obtener una mayor resolución que con las otras técnicas. Sin embargo, las moléculas de tamaño menor a aproximadamente 200 kDa suelen ser difíciles de estudiar mediante criomicroscopía debido al bajo contraste de esta técnica. La criotinción negativa es una técnica intermedia entre las dos anteriores, ya que consiste en vitrificar la muestra en una solución que contiene un agente de tinción para incrementar el contraste. Esta técnica combina el contraste proporcionado por la tinción negativa, a la vez que preserva la estructura nativa de la molécula mediante la vitrificación. Sin embargo, la resolución alcanzada suele ser del orden de la obtenida con tinción negativa (Adrian et al., 1998).

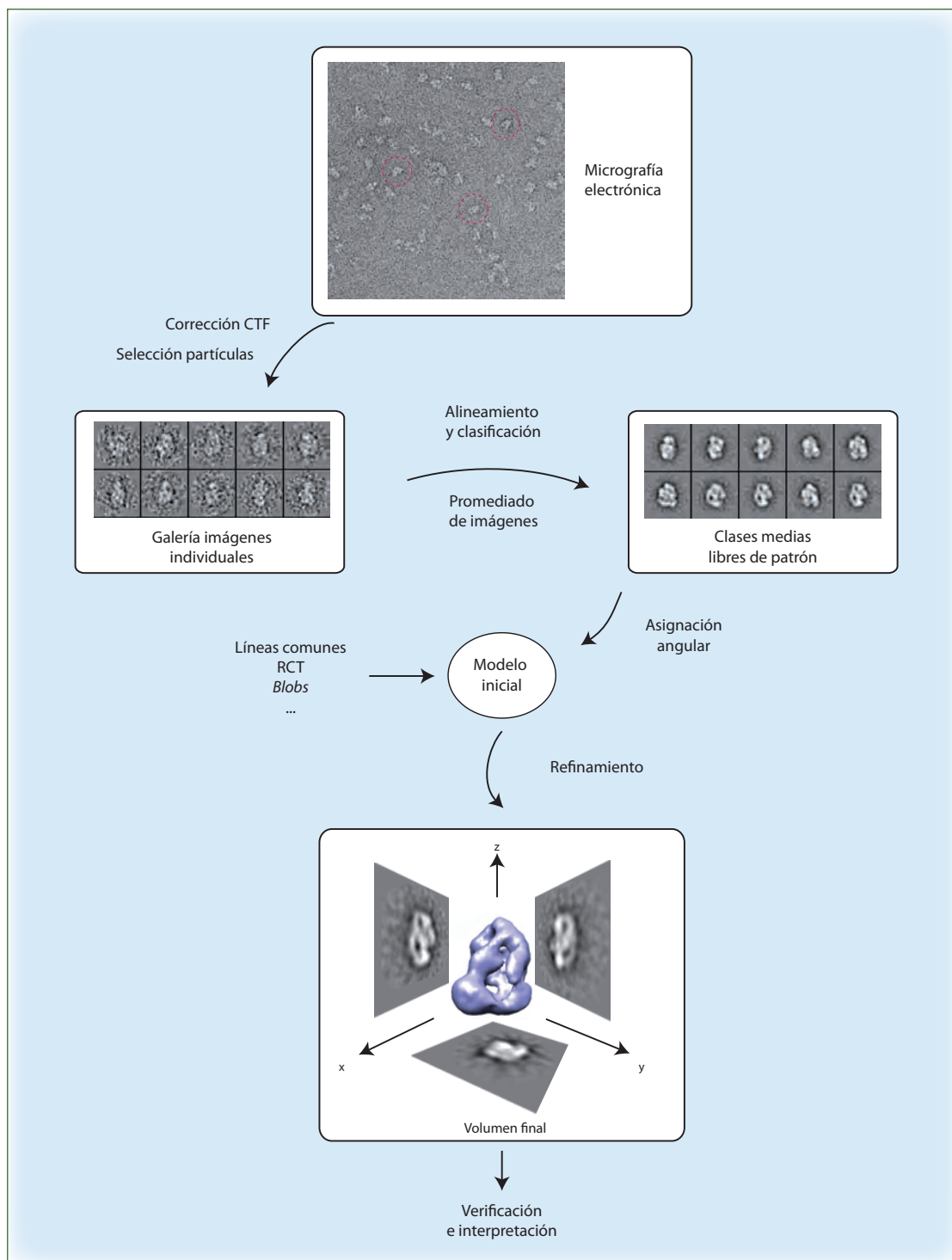
Después de colocar la rejilla en el microscopio, a la hora de la toma de datos es necesario hacerlo con una dosis mínima de electrones para que éstos no dañen la estructura biológica en exceso y se pueda preservar así la máxima información estructural posible. Esto genera un problema adicional, y es que las imágenes tomadas presentan una relación señal-ruido muy baja que condiciona el análisis posterior de las imágenes. Los electrones, al atravesar el espécimen interaccionan con él y llegan a un detector que recoge la información, y que puede ser una cámara CCD (*Charged Coupled Device*) o placas fotográficas (van Heel et al., 2000). En el primer caso, las imágenes se toman en formato digital, mientras que en el segundo se requiere un paso de digitalizado previo al procesamiento de imagen.

Una vez tomadas las micrografías necesarias, se procede a extraer miles de imágenes de moléculas individuales a partir de un fichero digital de las micrografías, y a realizar el procesamiento de imagen y análisis computacional para determinar la estructura tridimensional del objeto (Figura 1.10). El procesamiento de imágenes puede llevarse a cabo mediante distintos paquetes informáticos, cada uno de los cuales utiliza diferentes estrategias en cada uno de los pasos descritos anteriormente. Entre los más cono-

cidos y usados se encuentran EMAN (Ludtke et al., 1999), SPIDER (Frank et al., 1996), IMAGIC (van Heel et al., 1996) y XMIPP (Marabini et al., 1996; Sorzano et al., 2004). Estos paquetes contienen un conjunto de operaciones y comandos aplicables a los datos obtenidos mediante microscopía. Estos comandos presentan multitud de parámetros que necesitan ser controlados y supervisados por el usuario, de forma que todo el proceso de reconstrucción requiere la interacción y toma de decisiones del usuario. Es importante destacar que las imágenes tomadas en el microscopio electrónico son proyecciones en dos dimensiones de moléculas individuales del espécimen bajo estudio, las cuales recogen toda la información contenida en la molécula en esa dirección. Estas imágenes representan diferentes vistas del espécimen, siempre y cuando éste se haya dispuesto sobre la rejilla en diferentes orientaciones. Además, las imágenes tomadas en el microscopio electrónico están afectadas por un efecto conocido como función de transferencia de contraste o CTF (*Contrast Transfer Function*). Este efecto proviene de las aberraciones de las lentes del microscopio, y está modulado por el desenfoque utilizado para visualizar las imágenes, además de por la coherencia de los electrones emitidos. Para obtener la estructura de la molécula original es necesario determinar y corregir estos efectos, con el fin de poder recuperar la máxima información contenida en las imágenes de proyección.

Posteriormente, las imágenes correspondientes a vistas similares de las moléculas se alinean, clasifican y promedian sin utilizar más información que la contenida en los propios datos, con el objetivo de aumentar la señal y reducir el ruido producido por la baja dosis de electrones empleada en la adquisición de las imágenes. Técnicamente, este proceso se conoce como obtención de imágenes medias libres de patrón. Las diferentes clases obtenidas tras la clasificación representan las diferentes vistas de la molécula en diferentes orientaciones, siempre y cuando todas las moléculas provengan del mismo espécimen y en la misma conformación. Este análisis en dos dimensiones puede facilitar por sí solo información relevante sobre la estructura de la molécula y a veces puede ser utilizada como estrategia para localizar regiones dentro de un complejo de proteínas (Figura 1.10).

A continuación es necesario determinar la orientación relativa entre las medias obtenidas para reconstruir la estructura de la que provienen. Encontrar la relación angular correcta es uno de los pasos limitantes en microscopía electrónica, por lo que se pueden cometer errores en la asignación que pueden dar lugar a estructuras incorrectas. Si la molécula se ha dispuesto en todas las orientaciones posibles, existirá una buena cobertura de ángulos a la hora de realizar la reconstrucción tridimensional, la cual es necesaria para obtener la estructura correcta (Llorca, 2005). Existen diferentes estrategias para determinar la relación angular, pero todas pasan por la obtención de un modelo inicial. Éste puede ser generado a partir de los propios datos mediante el procedimiento denominado “líneas comunes” o bien de manera independiente a éstos. Este mapa inicial es luego refinado con el conjunto de datos en un proceso iterati-



**Figura 1.10. Representación esquemática del proceso de reconstrucción de partículas individuales mediante microscopía electrónica.** Las imágenes de proyección que se obtienen experimentalmente se relacionan con rotaciones específicas en los ejes X, Y y Z de la estructura tridimensional correspondiente.

vo hasta obtener un modelo final. Una vez obtenida la estructura, se procede al cálculo de resolución para determinar el grado de detalle y la información proporcionada por la estructura de la molécula. Existen diversas formas de determinar la resolución, pero

existe controversia al respecto porque no hay un criterio unificado para llevarlo a cabo (Rosenthal y Henderson, 2003; van Heel y Schatz, 2005).

Finalmente, todo análisis realizado mediante microscopía electrónica requiere una verificación e interpretación del volumen final obtenido, siendo éste otro de los pasos complicados en el proceso. Si existen estructuras atómicas resueltas mediante otras técnicas, se puede realizar un ajuste de las mismas en el mapa obtenido mediante microscopía. Si no existen estructuras disponibles, se puede llevar a cabo la localización de subunidades o dominios mediante marcaje con anticuerpos, oro, proteínas de fusión o péptidos, delección de dominios, entre otras estrategias.



## **OBJETIVOS**



El objetivo general de esta tesis doctoral es profundizar en el conocimiento estructural del proceso de reparación de roturas de la doble cadena del DNA mediante el mecanismo de unión de extremos no homólogos.

Para alcanzar este objetivo, se han planteado los siguientes objetivos concretos:

- Purificación y caracterización bioquímica de la proteína Xrcc4 y del complejo Xrcc4-DNA ligasa IV
- Determinación de la estructura de la proteína Xrcc4 mediante microscopía electrónica
- Caracterización estructural del complejo Xrcc4-DNA ligasa IV mediante microscopía electrónica
- Purificación de la proteína XLF y análisis de su interacción con el complejo Xrcc4-DNA ligasa IV





# **MATERIALES Y MÉTODOS**



### 3.1. Clonación, expresión y purificación de proteínas

Tanto la clonación como la purificación de las proteínas Xrcc4, así como del complejo Xrcc4-DNA ligasa IV y las proteínas XLF trucadas fueron llevadas a cabo como parte de la colaboración con el grupo del profesor Laurence Pearl en el Instituto de Investigación del Cáncer (*Institute of Cancer Research*), Londres. Todas ellas aparecen detalladas en la [Tabla 3.1](#).

#### 3.1.1. Clonación, expresión y purificación de Xrcc4 de *Mus musculus*

El cDNA de la proteína Xrcc4 de ratón se clonó en el plásmido pRSET-B (Invitrogen). Para la expresión de la proteína se utilizó la cepa *E.coli* BL21-CodonPlus (DE3) de Stratagene. La proteína Xrcc4 expresada contenía un polipéptido de seis residuos de histidinas en su extremo N-terminal y se purificó mediante cromatografía de afinidad sobre una columna de Ni-NTA (Clontech). A continuación se realizó una cromatografía de intercambio aniónico sobre una columna HiTrap Q (GE Healthcare). Por último, se llevó a cabo una cromatografía de exclusión molecular en una columna Superdex 200 (GE Healthcare). La proteína se obtuvo a una concentración final de 22  $\mu$ M en un tampón que contenía 20 mM Tris-HCl pH 8,0, 10% glicerol, 200 mM NaCl, 1 mM EDTA, 2 mM DTT y 0,1 mM PMSF.

#### 3.1.2. Clonación, expresión y purificación del complejo Xrcc4-DNA ligasa IV humano

Para la coexpresión de la proteína Xrcc4 y DNA ligasa IV se utilizó el plásmido pFastBac Dual (Invitrogen). El cDNA de la proteína Xrcc4 se clonó en el sitio de restricción de la enzima *StuI*. El cDNA de la proteína DNA ligasa IV fue clonado entre los sitios de restricción *BamHI* y *NotI*. Durante el clonaje se fusionó en fase el cDNA de la proteína con la secuencia correspondiente a seis residuos de histidinas, de modo que a partir de este plásmido se pudiera expresar la proteína His6-DNA ligasa IV. La expresión de las proteínas se llevó a cabo mediante la infección de células SF9 con baculovirus recombinantes derivados del plásmido anterior (sistema Bac-to-Bac de Invitrogen). Para realizar la purificación del complejo Xrcc4-His6-DNA ligasa IV las células se centrifugaron y se lisaron con un tampón que contenía 20 mM Tris-HCl pH 8,0, 250 mM NaCl, 5% glicerol, 1 mM DTT e inhibidores de proteasas sin EDTA (Roche). Para la purificación de la proteína presente en la fracción soluble, se realizó un primer paso de cromatografía de afinidad sobre una matriz de Ni-NTA (Clontech) equilibrada con el tampón de lisis. La resina se lavó con el mismo tampón suplementado con imidazol a una concentración de 5 mM, y la proteína retenida se eluyó posteriormente con un tampón que contenía 50 mM Tris-HCl pH 7,0, 50 mM NaCl, y 250 mM imidazol. Las fracciones que contenían el complejo se diluyeron con un tampón 30mM Tris-HCl pH 8,0, 50 mM NaCl, 2 mM DTT y se purificaron a través de una columna de intercambio aniónico HiTrap Q (GE

Healthcare) de 5 ml. El complejo se eluyó con un gradiente lineal de NaCl de 50 a 600 mM. A continuación, la proteína eluida se purificó utilizando una columna de heparina. Finalmente, se realizó una cromatografía de exclusión molecular empleando una columna Superdex 200 (GE Healthcare) que permitió purificar el complejo a homogeneidad. La concentración final del complejo purificado fue de 2,8  $\mu$ M, en un tampón que contenía 20 mM Tris-HCl pH 8,0, 200 mM NaCl, 1 mM DTT y 5% glicerol.

### 3.1.3. Clonación, expresión y purificación del complejo Xrcc4( $\Delta$ 214-336)-DNA ligasa IV .

El complejo constituido por una versión truncada de Xrcc4, Xrcc4( $\Delta$ 214-336), unida a DNA ligasa IV completa se purificó siguiendo métodos similares. Para facilitar la purificación de las proteínas se expresó Xrcc4 con una cola de seis histidinas en su extremo N-terminal. El complejo se purificó a una concentración de 39  $\mu$ M en un tampón 20 mM Hepes pH 7,0, 300 mM NaCl, 6 mM DTT y 30% glicerol.

### 3.1.4. Clonación, expresión y purificación de XLF humana, XLF( $\Delta$ 242-299) y XLF( $\Delta$ 248-299)

El cDNA de la proteína XLF silvestre (residuos 1-299), así como de las proteínas truncadas XLF( $\Delta$ 242-299) y XLF( $\Delta$ 248-299) se clonaron en el plásmido pGEX-6P1. Este plásmido permite la expresión de proteínas en fase con un polipéptido de GST en el extremo N-terminal. Para la expresión de las proteínas se utilizaron las cepas *E.coli* Rosetta2 y BL21(DE3). Las células se dejaron crecer hasta una OD600 de 0,6 y la expresión se indujo empleando distintas concentraciones de IPTG, desde 0,1 mM a 1 mM, utilizando distintas temperaturas (16°C a 37 °C) y distintos tiempos (desde 10 h a toda la noche). Se lisaron las células mediante sonicación en un tampón que contenía 20 mM Tris-HCl pH 7,0 y 200 mM NaCl. La proteína contenida en la fracción soluble se incubó con una resina de glutatión-sefarosa (GE Healthcare) y la proteína retenida se eluyó

Nombre	Origen	Residuos	P.Molecular
Xrcc4	Ratón	1-326	32 kDa
Xrcc4-DNA ligasa IV	Humano	1-336; 1-911	180 kDa
Xrcc4( $\Delta$ 214-336)-DNA ligasa IV	Humano	1-213;1-911	150 kDa
XLF	Humano	1-299	30 kDa
XLF( $\Delta$ 242-299)	Humano	1-241	24 kDa
XLF2( $\Delta$ 248-299)	Humano	1-247	24 kDa

**Tabla 3.1. Proteínas y complejos estudiados en esta tesis.** En la tabla se especifican los nombres utilizados para cada proteína o complejo, así como la fuente de purificación, el número de residuos que tienen y su peso molecular.

con glutatión y tras un proceso de diálisis se incubó con la proteasa 3C durante toda la noche. La identidad e integridad de las proteínas se confirmó mediante espectrometría de masas.

#### 3.1.5. Purificación de Xrcc4 mediante centrifugación preparativa en gradientes de glicerol

Se depositaron 10  $\mu$ l de la proteína Xrcc4 a una concentración de 22  $\mu$ M sobre un gradiente de glicerol discontinuo del 0% al 60%. Como control, una solución de diversos marcadores (BioRad) fueron analizados en paralelo. La centrifugación se llevó a cabo en una ultracentrífuga Beckman TL-100 equipada con un rotor TLS-55 durante 24 horas a 25000 rpm y 4°C. Se recogieron fracciones de 100  $\mu$ l desde la parte superior del gradiente y se analizó la presencia de la proteína Xrcc4 en las mismas mediante electroforesis SDS-PAGE.

#### 3.1.6. Purificación del complejo Xrcc4-DNA ligasa IV mediante cromatografías de exclusión molecular

El complejo Xrcc4-DNA ligasa IV se analizó mediante cromatografía de exclusión molecular. Se aplicó el complejo a una concentración 2,3  $\mu$ M sobre una columna Superdex 200 equilibrada con un tampón que contenía 20 mM HEPES pH 7,5, 200 mM NaCl, 5% glicerol.

### 3.2. Electroforesis en condiciones nativas y desnaturalizantes (SDS-PAGE)

El análisis de las proteínas se realizó mediante electroforesis en condiciones desnaturalizantes o nativas, utilizando geles del porcentaje adecuado en cada caso. El tampón empleado en las electroforesis en condiciones desnaturalizantes contenía Tris-HCl 25 mM, Glicina 190 mM y SDS 0,1%. Las muestras se cargaron en un tampón que contenía Tris-HCl 30 mM pH 6,8, SDS 1%, B-ME 5%, glicerol 10%, azul de bromofenol 0,25 mg/ml. El tampón empleado en las electroforesis en condiciones nativas estaba constituido por Tris-HCl 25 mM pH 8,0 y glicina 190 mM. La electroforesis en condiciones nativas se desarrolló en un tampón TBE constituido por Tris 90mM, ácido bórico 90 mM, EDTA 2,5 mM, pH 8,3.

Una vez realizada la electroforesis, los geles se revelaron de distintas formas:

**Tinción con Coomassie.** La solución de tinción contiene Coomassie Brilliant Blue R-250 (BioRad) 1mM, etanol 48% v/v y ácido acético 25% v/v. La solución de destinción está compuesta por ácido acético 10% y etanol 9,6%.

**Tinción con plata.** Los geles de poliacrilamida SDS-PAGE se revelaron de acuerdo

al siguiente protocolo: Incubación durante 30 min en una solución 40% etanol y 10% ácido acético. Después se incubó 30 min en una solución 30% etanol, 0,5% glutaraldehído, 0,2% tiosulfato sódico y 6,8% acetato sódico. Tras tres lavados de agua de 5 min, se incubaron durante 20 min en una solución con 0,1% nitrato de plata y 0,04% formaldehído. A continuación se revelaron en una solución 2,5% carbonato sódico y 0,02% formaldehído y por último en una solución de 1,5% EDTA durante 10 min.

### 3.3. Experimentos de ultracentrifugación analítica

La ultracentrifugación analítica puede usarse para realizar dos tipos principales de experimentos: velocidad de sedimentación y equilibrio de sedimentación. La velocidad de sedimentación es una técnica hidrodinámica que está relacionada con la masa y la forma de las macromoléculas que se analizan y que permite la obtención del valor del coeficiente de sedimentación característico de cada una de ellas. En los experimentos de velocidad de sedimentación se hacen medidas de la concentración de las macromoléculas en función de la distancia radial  $r$  (perfiles de sedimentación) a intervalos regulares de tiempo, para determinar la velocidad de movimiento (relacionada con el coeficiente de sedimentación) y el ensanchamiento del frente de macromoléculas en movimiento (relacionado con el coeficiente de difusión) en función del tiempo, cuando se aplica una fuerza centrífuga elevada. Este tipo de ensayos proporciona información acerca del número de especies macromoleculares que constituyen la muestra.

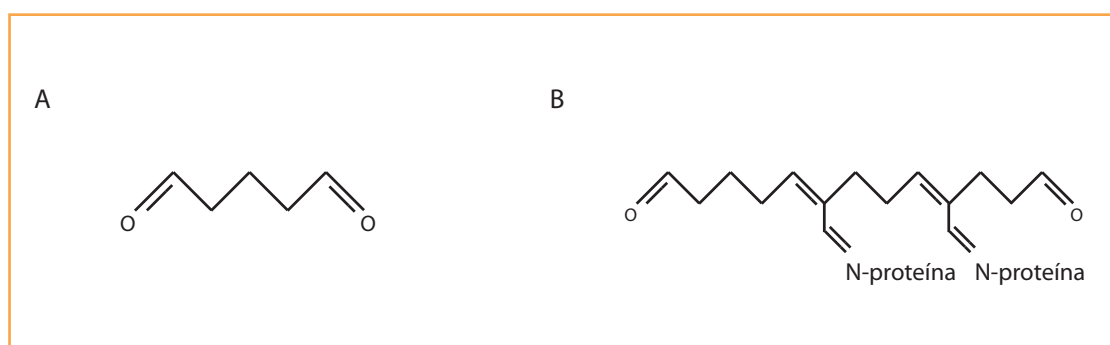
Los análisis mediante velocidad de sedimentación de Xrcc4 fueron realizados en el Servicio de Ultracentrifugación Analítica del Centro de Investigaciones Biológicas, Madrid. Para los ensayos se utilizaron 350  $\mu$ l de la muestra Xrcc4 a diferentes concentraciones (1,1, 2,2 y 4,4  $\mu$ M) en un tampón 20 mM Tris pH 8,0, 200 mM NaCl, 1% glicerol. Los experimentos se llevaron a cabo a 40000 rpm durante 2 horas a 20°C en una ultracentrífuga analítica Optima XL-A (Beckman) utilizando un rotor An50Ti. El análisis de los perfiles de sedimentación obtenidos se realizó mediante el ajuste directo de soluciones a la ecuación de Lamm a los perfiles con el programa SEDFIT (Schuck y Rossmannith, 2000). Este ajuste no emplea valores discretos de los coeficientes de sedimentación, sino distribuciones de coeficientes de sedimentación. El mejor ajuste obtenido de los perfiles de sedimentación experimentales a las soluciones de la ecuación de Lamm proporciona una distribución  $c(s)$  que representa la concentración inicial de especies que sedimentan con un coeficiente de sedimentación  $s$ . El coeficiente de sedimentación se corrigió con el programa SEDNTERP a condiciones estándar  $S_{20,w}$  (20 °C en agua).

### 3.4. Ensayos de entrecruzamiento con glutaraldehído

El entrecruzamiento o *cross-linking* es un proceso químico de unión covalente que ocurre entre dos o más moléculas. Los agentes de entrecruzamiento contienen en sus extremos grupos reactivos frente a determinados grupos funcionales (aminas primarias, entre otras) que se encuentran en moléculas como las proteínas. Estos compuestos se utilizan para determinar interacciones entre moléculas cercanas con el fin de determinar la estructura tridimensional de complejos, entre otras muchas aplicaciones. En la mayoría de los casos es necesario mantener la estructura nativa del complejo de proteínas, de forma que los ensayos de entrecruzamiento se realizan en condiciones de pH y tampón suaves. Además, la relación molar óptima entre agente de entrecruzamiento y proteína es un factor importante, además del tiempo de entrecruzamiento, puesto que tiene que haber un equilibrio entre entrecruzamiento y agregación. Como agente de entrecruzamiento se empleó glutaraldehído (SIGMA) (Figura 3.1, A), el cual lleva a cabo un entrecruzamiento directo de sus grupos amino con los grupos amino de las proteínas (Figura 3.1, B).

Se llevaron a cabo dos tipos de experimentos de entrecruzamiento. Por un lado, se incubó Xrcc4 a una concentración final de 4  $\mu\text{M}$  con glutaraldehído al 0,06%. La reacción se llevó a cabo a temperatura ambiente y durante distintos tiempos de incubación desde 1 a 15 min en tampón 20 mM Hepes pH 7,0, 200 mM KCl, 0,1 mM DTT, en un volumen de reacción de 12  $\mu\text{l}$ . En un segundo experimento se incubó Xrcc4 (4  $\mu\text{M}$ ) con concentraciones crecientes de glutaraldehído del 0,02% al 0,1% y durante 5 min. Las reacciones se detuvieron añadiendo 2  $\mu\text{l}$  de un tampón que contenía Tris-HCl 1 M pH 7,5. Los productos de la reacción se analizaron mediante electroforesis en condiciones desnaturalizantes en geles al 6%.

Los experimentos de entrecruzamiento del complejo Xrcc4-DNA ligasa IV se llevaron a cabo en condiciones similares a las descritas anteriormente para la proteína Xrcc4.



**Figura 3.1. A)** Estructura de una molécula de glutaraldehído. **B)** Mecanismo de entrecruzamiento con proteínas llevado a cabo por el glutaraldehído. Los dos grupos reactivos aldehído le permiten establecer enlaces cruzados entre diversos tipos de moléculas, formando uniones covalentes entre ellos.



En este caso, el complejo se utilizó a una concentración de 2,3  $\mu\text{M}$  en un volumen final de 10  $\mu\text{l}$ .

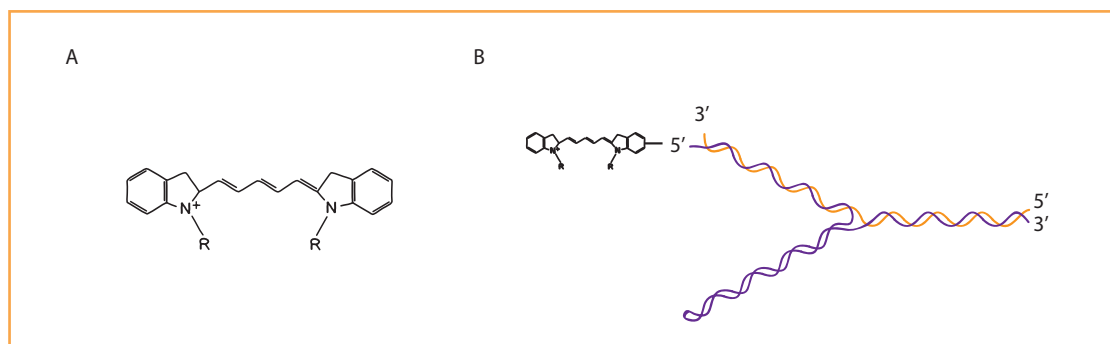
### 3.5. Caracterización de los complejos entrecruzados mediante espectrometría de masas

Las bandas correspondientes a Xrcc4 y DNA ligasa IV libres, así como a las tratadas con glutaraldehído se extrajeron del gel y tras un procesamiento previo de las mismas, se sometieron a un proceso de digestión con tripsina, se analizó la composición peptídica y se identificaron las proteínas mediante espectrometría de masas MALDI-TOF en la Unidad de Proteómica del Parque Científico de Madrid. La asignación de las masas específicas de los péptidos representados en los espectros obtenidos se realizó mediante el programa MASCOT ([http://www.matrixscience.com/search\\_form\\_select.html](http://www.matrixscience.com/search_form_select.html)).

### 3.6. Ensayos de retardo en la movilidad electroforética en complejos proteína-DNA

Los oligonucleótidos empleados en los ensayos de retardo en gel o ensayos de retardo en la movilidad electroforética EMSA (*Electrophoretic Mobility Shift Assays*) eran de dos tipos. Por un lado, los oligonucleótidos utilizados en la unión a Xrcc4, al complejo Xrcc4-DNA ligasa IV y XLF eran fragmentos de DNA de 390 pb de doble cadena originados a partir del fragmento de DNA mLST8pET15b mediante amplificación por PCR, el cual se obtuvo a una concentración de 1,77  $\mu\text{M}$ . Las reacciones de unión se llevaron a cabo en un tampón 20 mM Hepes pH 7,9, 2 mM EDTA, 2 mM EGTA, 5 mM DTT, 37,5 mM KCl, 5% glicerol durante 1 hora a temperatura ambiente en un volumen de reacción de 10  $\mu\text{l}$ . Después se llevó a cabo un ensayo de retardo en gel de poliacrilamida al 5 % en condiciones nativas con el fin de determinar si se habían formado los complejos con DNA. Como control se analizaron las mismas cantidades de proteína y de DNA independientemente en el mismo gel. Estos geles fueron teñidos con bromuro de etidio para detectar el DNA, y con una solución de plata para detectar las proteínas.

Por otro lado, en los ensayos de EMSA con el complejo Xrcc4-DNA ligasa IV se diseñaron unos oligonucleótidos siguiendo el protocolo descrito anteriormente (Spagnolo et al., 2006). Los oligonucleótidos directos presentan una horquilla que impide que las proteínas se deslicen a lo largo de la molécula de DNA y además llevan un fluoróforo incorporado en el extremo 5'-P, denominado Cy5 para que puedan ser detectados mediante fluorescencia. El Cy5 es un fluoróforo que tiene un máximo de absorbancia a 643 nm y un máximo de emisión a 667 nm y es incorporado como una amidita durante la síntesis de los oligos. Los oligos fueron sintetizados por Sigma (Figura 3.2). La se-



**Figura 3.2. A)** Estructura de una molécula de Cy5. **B)** Estructura de la molécula de DNA de doble cadena generada tras la hibridación de los oligos. La molécula Cy5 está acoplada al extremo 5'-P de la cadena de DNA que presenta la horquilla o del oligo delantero.

cuencia de los oligonucleótidos se detalla en la [Tabla 3.2](#). Los oligonucleótidos fueron hibridados para generar un DNA de doble cadena de 35 pares de bases con extremos romos.

En cada reacción de anillamiento se incubó el oligonucleótido marcado con Cy5 a una concentración de 0,5  $\mu\text{M}$  con un exceso del oligonucleótido no marcado (1,5  $\mu\text{M}$ ), en un tampón que contenía 10 mM Tris-HCl pH 8,0, 1 mM EDTA y 100 mM NaCl. La mezcla se incubó a 95  $^{\circ}\text{C}$  durante 10 min en un baño con agua y se dejó enfriar posteriormente hasta alcanzar la temperatura ambiente. El producto de la reacción se analizó en un gel nativo de acrilamida al 16% para comprobar si los oligonucleótidos habían hibridado.

En los experimentos de EMSA se incubó el DNA de doble cadena con el complejo Xrcc4-DNA ligasa IV en un tampón 20 mM HEPES pH 8,0, 2 mM EDTA, 1 mM DTT, 300 mM KCl y 5% glicerol. El volumen final de la reacción fue de 10  $\mu\text{l}$  y las muestras se incubaron durante 1 hora a temperatura ambiente. Los complejos se analizaron en geles de poliacrilamida al 6% o bien en geles de gradiente del 4-15% (BioRad), en ambos casos en condiciones nativas. En todos los casos, la electroforesis se llevó a cabo a 4  $^{\circ}\text{C}$  durante 2 horas en un tampón TBE al 0,5X. Para detectar el DNA se expuso el gel en un lector de fluorescencia Phosphorimager.

Nombre	Tamaño (nt)	Secuencia 5'-3'	Modificación
35_bp_bl_Fw	51	CCCAGCTTTCCAGCTAATAAACTA AAAAC TATTATTATGGCCGCACGCGT	5'(Cy5)
35bp_bl_Rv	38	A C G C G T G C G G C C A T A A T A A T A G T T T T T A G T T T A T T G G G	

**Tabla 3.2** Oligonucleótidos utilizados en los ensayos de interacción Xrcc4-DNA ligasa IV con DNA

### 3.7. Ensayos de interacción entre el complejo Xrcc4-DNA ligasa IV y la proteína XLF

Se incubó el complejo truncado Xrcc4( $\Delta$ 214-336)-DNA ligasa IV con una de las proteínas XLF que habían sido purificadas truncadas (XLF( $\Delta$ 248-299)) en una relación molar 1:3 durante una hora a temperatura ambiente en un tampón que contenía 200 mM NaCl, 20 mM Hepes pH 7,5, 5% glicerol y 5 mM DTT. A continuación se analizó la posible formación del complejo ternario mediante cromatografía de exclusión molecular en una columna Superdex 200. Se inyectaron en la columna como control del experimento las mismas cantidades de complejo Xrcc4-DNA ligasa IV de manera independiente. A continuación se analizaron las fracciones recogidas mediante electroforesis en condiciones desnaturalizantes.

### 3.8. Análisis estructural mediante microscopía electrónica y procesamiento de imágenes

Los complejos citados anteriormente se purificaron mediante columnas de filtración molecular o gradientes de glicerol previamente a su preparación para el análisis mediante microscopía electrónica de transmisión (TEM).

#### 3.8.1. Preparación de muestras mediante tinción negativa.

Las rejillas empleadas como soporte para la muestra eran de cobre de 400 agujeros, de las casas Aname y Fedelco. Se sometieron las rejillas a descarga iónica durante 1 min, con el fin de obtener una superficie hidrofílica que facilite la adsorción de la muestra. A continuación, la muestra (aproximadamente 4  $\mu$ l a una concentración entre 0,1  $\mu$ M y 1  $\mu$ M) se depositó sobre las rejillas durante 2 min. Después de lavarse con agua miliQ se tiñeron con acetato de uranilo al 2% p/v en unas ocasiones, o con formiato de uranilo al 2% p/v en otras, durante 1 min y se dejaron secar. Las rejillas se visualizaron en un microscopio electrónico de transmisión JEOL 1230, equipado con un filamento de wolframio y estabilizado a 100kV.

#### 3.8.2. Adquisición y digitalizado de imágenes.

Las micrografías electrónicas se tomaron a una magnificación de 40000 aumentos en condiciones de mínima dosis de electrones (en torno a  $10e^-/\text{\AA}^2$ ) sobre placas fotográficas Kodak SO-163. Las imágenes fueron digitalizadas utilizando un escáner MINOLTA Dimage Scan Multi Pro con una resolución de 2,6  $\text{\AA}$ /píxel.

#### 3.8.3. Procesamiento de imágenes y reconstrucción tridimensional.

La función de transferencia de contraste (*Contrast Transfer Function*, CTF) fue es-

timada para cada micrografía mediante el programa CTFFIND3 (Mindell y Grigorieff, 2003) y corregida utilizando el programa BSOFT (Heymann y Belnap, 2007). A continuación, las imágenes correspondientes a partículas individuales se extrajeron utilizando un conjunto de programas incluidos en la plataforma EMAN (Ludtke et al., 1999), se promediaron hasta una resolución final de 5,2 Å/px, y se normalizaron y filtraron con distintos algoritmos disponibles en los paquetes informáticos EMAN y XMIPP (Marabini et al., 1996; Sorzano et al., 2004). Se realizó un análisis inicial de los datos en dos dimensiones mediante los programas refine2d.py de EMAN y ML2D de XMIPP (Scheres et al., 2005), algoritmos que llevan a cabo clasificaciones y alineamientos de las imágenes libres de patrón. En el primer caso, la clasificación está basada en correlación cruzada, mientras que en el segundo caso utiliza un algoritmo de máxima verosimilitud.

Para el refinado tridimensional de los datos se utilizaron diferentes estrategias en la obtención los volúmenes iniciales. En una de ellas se seleccionaron imágenes medias obtenidas mediante clasificaciones libres de referencia con las que se generó un volumen inicial mediante líneas comunes. En un abordaje diferente, se generaron de manera artificial elipsoides con una distribución de masa gaussiana, los cuales fueron utilizados como volúmenes iniciales en el refinamiento. En otras ocasiones, se emplearon como volúmenes de referencia estructuras atómicas resueltas previamente mediante cristalografía de rayos X, con el objeto de comprobar que los resultados obtenidos fueran comparables a los resultados obtenidos mediante los procesamientos mencionados anteriormente.

El refinamiento de los volúmenes se realizó mediante un proceso de refinamiento angular iterativo que fue llevado a cabo mediante el programa refine de EMAN inicialmente, y posteriores ciclos de refinado mediante el programa ML3D de XMIPP (Scheres et al., 2007), basado en máxima verosimilitud. En el caso de Xrcc4 se asumió simetría C2 (simetría rotacional de orden dos alrededor del eje longitudinal de la molécula) durante todo el proceso de refinamiento, y sólo se utilizaron en la reconstrucción las proyecciones de la molécula que podrían corresponder a vistas laterales de la misma, fácilmente identificables debido a la forma alargada de la proteína.

#### 3.8.4. Cálculo de la resolución y visualizado de los volúmenes

El procesamiento finalizó cuando las diferencias entre volúmenes de ciclos sucesivos se minimizaron. Para determinar la resolución de la reconstrucción obtenida, la población de partículas asignadas a cada una de las clases en la última iteración del proceso de refinamiento se dividió en dos grupos de forma aleatoria, las cuales fueron utilizadas para reconstruir volúmenes independientes. Estas dos reconstrucciones se utilizaron para estimar la resolución, determinando el grado de similitud entre ambos volúmenes mediante correlación en función de la frecuencia espacial. Para la estimación de la resolución se fijó un valor de corte de 0,5 en el coeficiente de correlación

en el espacio de Fourier o *Fourier Shell Correlation* (FSC). El programa utilizado para la visualización de los mapas de microscopía y las estructuras atómicas fue UCSF Chimera (Pettersen et al., 2004).

### 3.9. Predicción de estructura secundaria y alineamientos de secuencias

La secuencia de Xrcc4 fue extraída de la base de datos ExPASy Molecular Biology Server (Gasteiger et al., 2003). La predicción de estructura secundaria se llevó a cabo utilizando el servidor Phyre (Kelley y Sternberg, 2009). Los alineamientos de secuencia se realizaron con el programa ClustalW (Larkin et al., 2007) y la visualización de los alineamientos se realizó con el editor de alineamientos múltiples JalView (Waterhouse et al., 2009).

### 3.10. Modelado de proteínas y ajuste en la reconstrucción tridimensional obtenida mediante microscopía electrónica

La predicción de la estructura tridimensional de proteínas se lleva a cabo cuando no existe un modelo atómico conocido de la misma. Existe un gran número de servidores y programas que llevan a cabo alineamientos de secuencias de aminoácidos entre dos o más proteínas con el fin de encontrar homólogos, o predicen zonas de estructura secundaria en las secuencias. También pueden predecir la estructura tridimensional de proteínas, lo cual llevan a cabo de dos formas: por homología con otras proteínas de estructura conocida o mediante plegamiento *de novo* si no existe homología con otras proteínas.

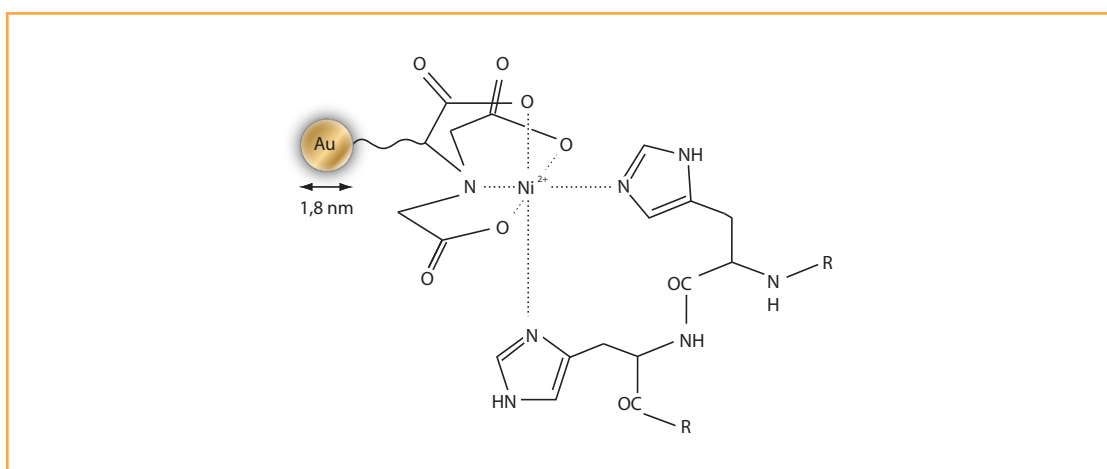
Se utilizó el servidor Robetta con el objetivo de encontrar homólogos estructurales a Xrcc4 y obtener posibles modelos estructurales para esta proteína (Chivian et al., 2003; Chivian et al., 2005; Kim et al., 2004). Robetta es un servidor que engloba multitud de programas y que proporciona predicciones de estructura mediante modelos comparativos y *ab initio* de una manera completamente automatizada, utilizando la secuencia de aminoácidos de la proteína de interés. Los modelos comparativos obtenidos, diez por defecto, son originados a partir de PDBs homólogos detectados por los programas UW-PDB-BLAST o 3DJury-A1 y alineados con el método K\*SYNC (Chivian y Baker, 2006). Las regiones de conexión son ensambladas a partir de fragmentos y optimizadas para que se ajusten a la parte de la estructura que ya ha sido alineada (Rohl et al., 2004). Los dominios sin un homólogo en la base de datos PDB son modelados mediante el protocolo de inserción de fragmento *de novo* de Rosetta (Bonneau et al., 2002).

Los modelos obtenidos por el servidor Robetta se alinearon con la estructura atómica de Xrcc4 para generar homodímeros. Estos dímeros se utilizaron para realizar

experimentos de ajuste en el mapa final de microscopía obtenido mediante el procesamiento de imágenes. El proceso de ajuste consiste en, bien de una forma manual o bien automática, encajar una estructura o modelo a resolución atómica, en un mapa de densidad obtenido mediante microscopía electrónica, el cual suele tener una menor resolución. Todos los ajustes fueron llevados a cabo con el programa informático ADP\_EM (Garzon et al., 2007), el cual utiliza un método de búsqueda exhaustiva en 6 dimensiones, explorando todas las posibles translaciones y rotaciones de la estructura atómica dentro del mapa de microscopía. En este caso, se utilizó un método de correlación cruzada, en el cual se tiene en cuenta toda la densidad de ambas estructuras en el momento de hacer el ajuste.

### 3.11. Marcaje de proteínas mediante moléculas de oro unidas a níquel

Los complejos Xrcc4-DNA ligasa IV, en los cuales la ligasa contenía seis histidinas en su extremo N-terminal, fueron incubados con un exceso molar (relación 1:4) de Ni-NTA-Nanogold (Nanoprobos) (Hainfeld et al., 1999), constituido por una molécula de oro de 1,8 nm de diámetro unida a una de ácido nitriloacético y a un grupo de níquel (Figura 3.3). Las proteínas purificadas mediante un polipéptido de histidinas tienen una afinidad extremadamente alta por determinados iones metálicos, por lo que estas moléculas pueden ser utilizadas para localizar dominios proteicos mediante microscopía electrónica. La muestra fue incubada con Ni-NTA-Nanogold durante 45 min a temperatura ambiente y 300 rpm. Los conjugados fueron purificados del exceso de partículas de oro no unidas mediante una cromatografía de exclusión molecular en una columna Superdex 200 equilibrada en un tampón que contenía 20 mM Tris-HCl pH 8,0, 200 mM NaCl y 25% glicerol. La presencia del complejo fue detectada mediante su absorbancia a 280 nm (refleja absorbancia de oro y proteína) y a 420 nm (absorbancia



**Figura 3.3. Estructura de una molécula de Ni-NTA Nanogold.** El Ni-NTA Nanogold está formado por una partícula de oro de 1,8 nm unido a un grupo de ácido nitriloacético y níquel, el cual interacciona con los grupos de histidinas de las proteínas.

de oro exclusivamente). La absorbancia a 420 nm se puede utilizar para calcular la parte de la absorbancia a 280 nm que es debida al Ni-NTA-Nanogold. Además se llevaron a cabo controles mediante cromatografías de filtración en gel en las que se inyectó el complejo o el oro de manera independiente, y también el complejo en presencia de 300 mM de imidazol tanto durante la incubación como en el tampón de equilibrio de la columna. Los conjugados recién purificados fueron aplicados sobre rejillas de cobre cubiertas con carbón (de las casas comerciales Aname y Fedelco) y visualizadas en el microscopio electrónico después de teñirlas negativamente con formiato de uranilo al 2%, siguiendo el protocolo descrito anteriormente. Tras la toma de micrografías, se llevó a cabo un análisis bidimensional con el fin de generar clases medias libres de patrón.

### 3.12. GraFix

GraFix es un método de purificación de complejos inestables desarrollado recientemente para su aplicación en la estabilización de complejos macromoleculares compatible con su observación y análisis mediante técnicas de microscopía electrónica (Kastner et al., 2008). Consiste en someter a la muestra a centrifugación en un gradiente de densidad, como glicerol, sacarosa o trehalosa, combinado con un agente de fijación débil, como puede ser el glutaraldehído. Esto da lugar a complejos monodispersos estabilizados químicamente que pueden ser analizados mediante microscopía electrónica, para realizar tanto tinción negativa como criomicroscopía. Este método fue aplicado al complejo Xrcc4-DNA ligasa IV, el cual una vez purificado fue cargado en un gradiente de glicerol discontinuo del 5% al 30% en un tampón constituido por 20 mM HEPES pH 7,4, 300 mM NaCl, 1 mM DTT, añadiendo 0,1% de glutaraldehído en la fracción que contenía mayor concentración de glicerol. La cantidad de muestra cargada fueron 10 µg a una concentración de 2,8 µM. La separación se llevó a cabo mediante centrifugación durante 18 horas a 25000 rpm y 4°C en una ultracentrífuga Beckman TL-100, empleando para ello un rotor TLS-55. En la primera fracción se cargó un colchón de 100 µl del mismo tampón pero sin glutaraldehído, ya que el tampón de la muestra contenía Tris-HCl y éste interfiere en el proceso. Las fracciones fueron recogidas desde la parte superior en volúmenes de 100 µl y analizadas en geles SDS-PAGE al 6%. Las fracciones que estaban enriquecidas en el complejo entrecruzado y que tenían un peso molecular aparente de 180 kDa, fueron adsorbidas sobre rejillas de cobre comerciales y analizadas mediante microscopía electrónica utilizando métodos similares a aquellos descritos anteriormente.





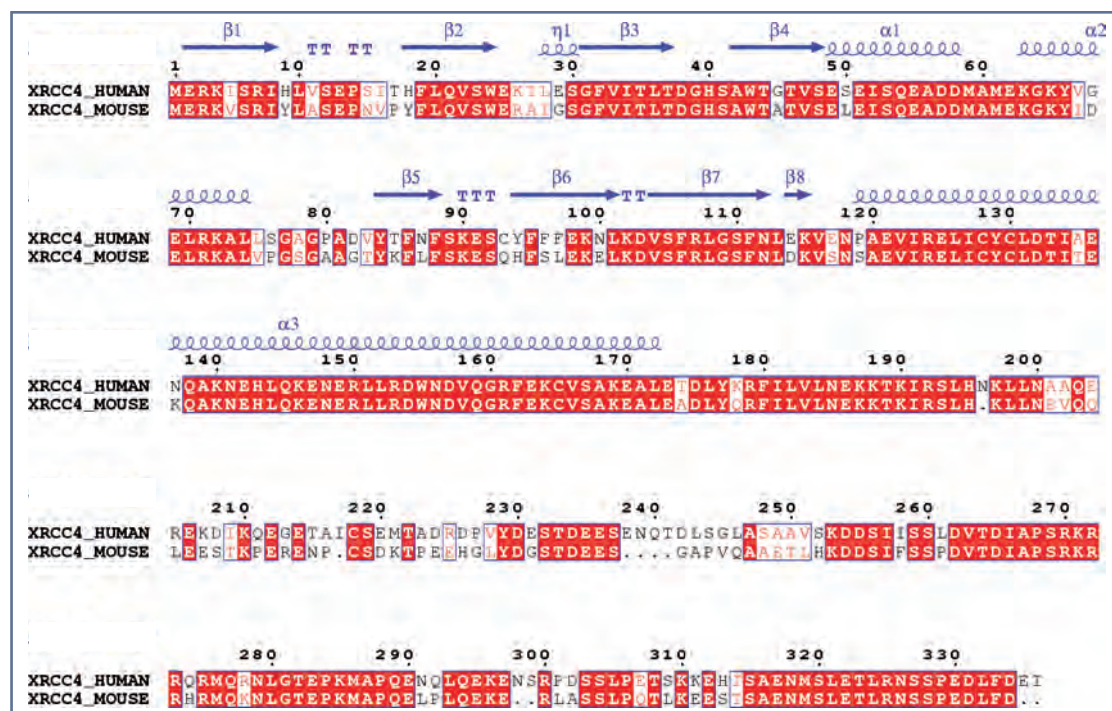


## RESULTADOS

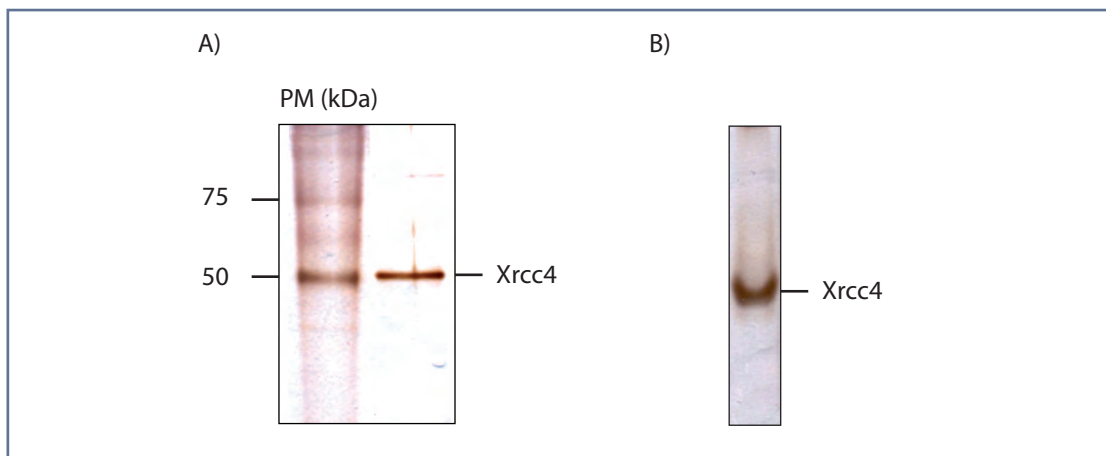


#### 4.1. Purificación de Xrcc4 de ratón

La proteína Xrcc4 fue purificada en colaboración con el grupo del Profesor Laurence Pearl en Londres (ver detalles en apartado 3.1.1 de Materiales y Métodos). Inicialmente, se intentó purificar la proteína Xrcc4 de origen humano, pero sin resultados positivos ya que ésta formaba oligómeros en geles nativos debido a la formación de puentes disulfuro intermoleculares. Por este motivo, se expresó y purificó Xrcc4 de ratón, la cual está constituida por 326 aminoácidos, diez residuos menos que la proteína humana (Critchlow et al., 1997), y no estaba afectada por este problema. Se llevó a cabo un alineamiento de las secuencias de ambas proteínas donde se determinó que estas proteínas presentan una elevada homología (76%), principalmente a nivel de la región N-terminal y C-terminal (Figura 4.1). La zona en la que presentan mayores diferencias es en la última región del dominio *coiled-coil* y parte del dominio C-terminal. Una vez expresada y purificada, la proteína se analizó mediante electroforesis en condiciones desnaturalizantes, en la que se pudo apreciar que Xrcc4 presenta una movilidad electroforética con un peso molecular aparente de 50 kDa (Figura 4.2, A), un comportamiento anómalo que ha sido observado de forma consistente por otros



**Figura 4.1. Alineamiento de las secuencias de Xrcc4 de ratón y humana.** Las secuencias fueron alineadas con el programa ClustalW, mediante el método del alineamiento entre pares. El grado de identidad de secuencia de los residuos de los motivos conservados se indican de la siguiente manera: el fondo rojo indica que los residuos son idénticos en ambas secuencias, mientras que el fondo blanco con los residuos en rojo muestra que las sustituciones son conservadas. También se muestran los motivos de estructura secundaria conocidos hasta el momento en Xrcc4 humana, cuya estructura parcial se ha resuelto mediante cristalografía de rayos X (extraído del archivo PDB: 1FU1).



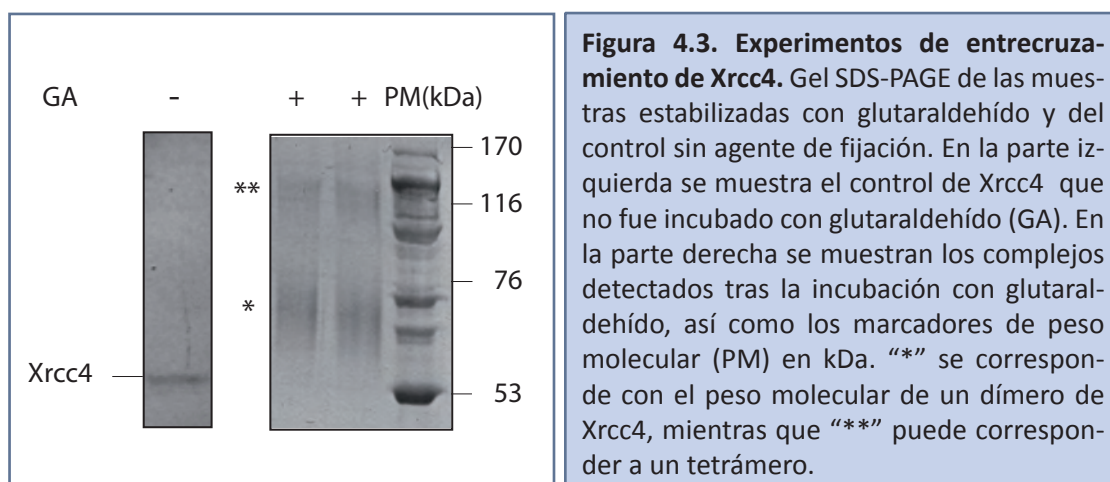
**Figura 4.2. Purificación de Xrcc4 previa a su análisis mediante microscopía electrónica. A)** Gel de poliacrilamida en condiciones desnaturalizantes de Xrcc4 teñido con plata. Los marcadores de peso molecular (PM) se muestran a la izquierda en kDa. **B)** Gel de poliacrilamida en condiciones nativas de Xrcc4.

grupos (Critchlow et al., 1997). También se llevó a cabo una electroforesis en condiciones nativas, en la que se comprobó que Xrcc4 migraba principalmente como una sola especie (Figura 4.2, B).

El análisis de Xrcc4 purificada mediante geles SDS-PAGE reveló que la proteína poseía un grado de pureza suficiente para iniciar su estudio estructural mediante microscopía electrónica. Sin embargo, al comenzar su análisis en el microscopio electrónico observamos que la muestra no era homogénea y que presentaba una elevada tendencia a la agregación. Además, se ha determinado previamente por otros grupos que Xrcc4 purificada a partir de fuentes endógenas ó recombinantes puede constituir tanto dímeros como tetrámeros. Mientras que las formas diméricas se ha observado que forman un complejo estable con la proteína DNA ligasa IV, la funcionalidad *in vivo* de las formas tetraméricas se desconoce en la actualidad (Junop et al., 2000; Leber et al., 1998). Por este motivo, decidimos llevar a cabo un análisis de la muestra mediante ensayos de entrecruzamiento y ultracentrifugación analítica con el fin de conocer el estado oligomérico de la muestra que habíamos purificado.

#### 4.2. Ensayos de entrecruzamiento de Xrcc4 con glutaraldehído

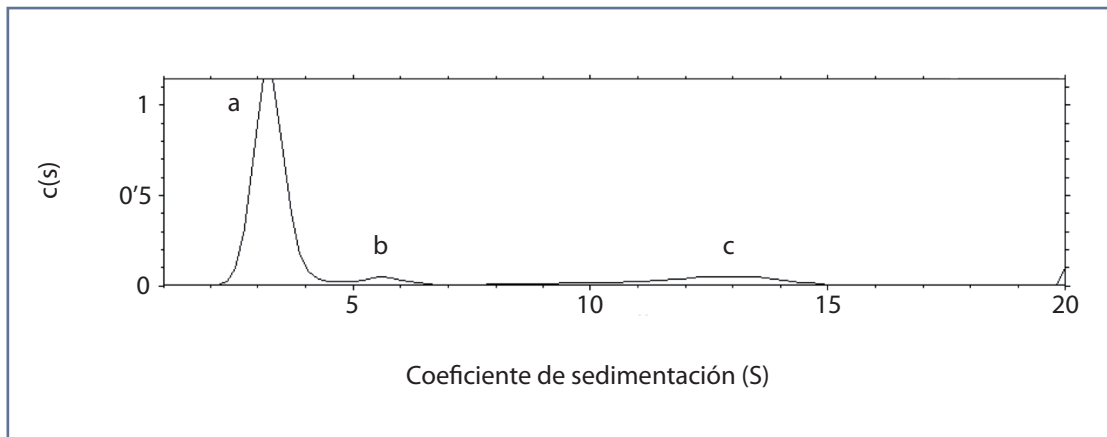
El glutaraldehído es un reactivo homobifuncional que reacciona con grupos amino primarios libres, como son los grupos  $\alpha$ -amino presentes en el extremo amino de las proteínas y los  $\epsilon$ -amino de los residuos lisina, y por este motivo, se analizó la capacidad del glutaraldehído de estabilizar los oligómeros de Xrcc4. Para ello, la proteína Xrcc4 purificada se incubó con diferentes concentraciones de glutaraldehído a distintos tiempos. A continuación, los complejos se resolvieron mediante electroforesis en geles de



poliacrilamida desnaturalizantes. Como control se analizó la muestra sin glutaraldehído. En la muestra incubada con fijador se observó una banda mayoritaria con una migración electroforética que, comparándola con los marcadores de peso molecular utilizados, correspondía a una proteína de alrededor de 76 kDa, compatible con un dímero de Xrcc4 (Figura 4.3 “\*”). Además, también fueron detectados otros componentes de mayor peso molecular, en concreto, se observó una especie que presentaba una movilidad electroforética de 170 kDa (Figura 4.3 “\*\*”) que podía corresponder con un tetrámero de Xrcc4. También se intuían bandas de mayor peso molecular que no se resolvieron en el gel y que correspondían probablemente a agregados. No se observaron diferencias significativas en los resultados obtenidos entre las distintas concentraciones de agente de fijación utilizadas ni tampoco entre los diferentes tiempos de incubación. Por tanto, los experimentos de entrecruzamiento llevados a cabo determinaron que los dímeros de Xrcc4 son la forma mayoritaria en nuestra preparación.

### 4.3. Análisis de Xrcc4 mediante ultracentrifugación analítica

Con el fin de confirmar el peso molecular de Xrcc4 determinado en los experimentos de entrecruzamiento, realizamos un análisis mediante velocidad de sedimentación que nos proporcionó información acerca de la homogeneidad, así como de la forma de las moléculas en solución. El perfil de velocidad de sedimentación de Xrcc4 (Figura 4.4) mostró que alrededor de un 80% de Xrcc4 sedimentaba como una única especie molecular con un coeficiente de sedimentación  $S_{20,w}$  de 3,2 S (Figura 4.4, pico a), cuyo valor es compatible con un dímero de Xrcc4. Acompañando a esta especie mayoritaria, aparecían dos especies de mayor peso molecular, con coeficientes de sedimentación de 5,80 S, lo cual era compatible con cuatro moléculas de Xrcc4 y que representaba un 5,5% del total de la muestra (Figura 4.4, pico b), y 13 S, el 12% restante, que corresponde con una alta probabilidad a agregados (Figura 4.4, pico c). El valor de la relación friccional  $f/f_0$  de la especie mayoritaria, calculado a partir del peso molecular de la molécula y el coeficiente de sedimentación calculado experimentalmente, fue de



**Figura 4.4. Experimentos de ultracentrifugación analítica de la proteína Xrcc4.** Se representa la distribución continua de concentraciones frente al coeficiente de sedimentación tras el análisis realizado mediante velocidad de sedimentación. Xrcc4 aparece principalmente como un dímero (a), aunque existen otros componentes de mayor peso molecular que pueden corresponder a tetrámeros de la proteína (b), así como a agregados (c).

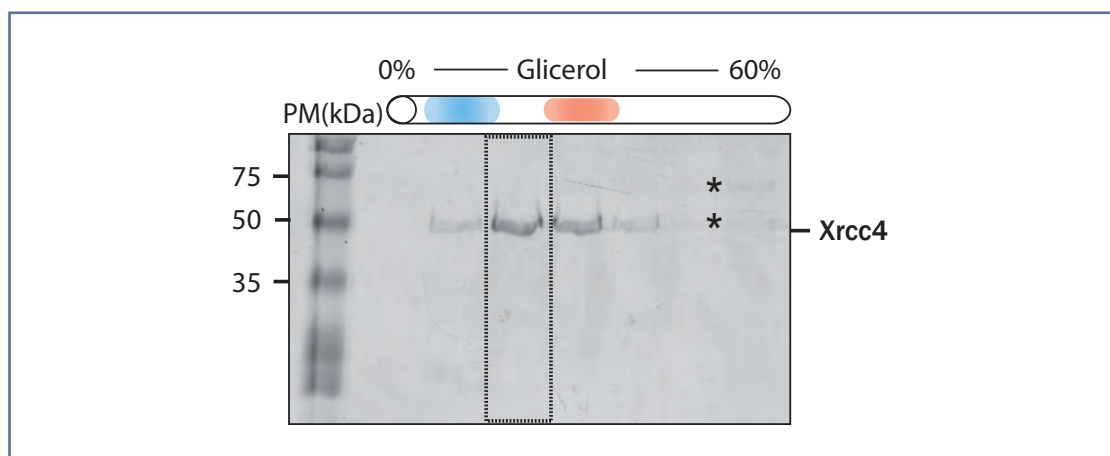
1,89. Este valor se encuentra en el rango determinado para proteínas moderadamente alargadas (1,6 a 1,9) (Erickson, 2009), lo cual concuerda con las estructuras resueltas previamente de determinados fragmentos de Xrcc4 (Dore et al., 2006; Junop et al., 2000; Sibanda et al., 2001; Wu et al., 2009). En estas condiciones, si la molécula fuera un monómero, sería completamente esférica.

#### 4.4. Purificación de los complejos diméricos de Xrcc4 mediante gradientes de glicerol previo a su análisis mediante microscopía electrónica

Con el fin de separar los agregados de los complejos diméricos y estabilizar la muestra previamente a su observación en el microscopio electrónico, Xrcc4 se sometió a un proceso de centrifugación en un gradiente discontinuo de glicerol. Las fracciones resultantes del gradiente se analizaron mediante SDS-PAGE (Figura 4.5), donde se observó que Xrcc4 se encontraba presente en distintas fracciones. En la figura se puede observar que en el fondo del gradiente aparecieron agregados de un alto peso molecular (marcados en la figura con “\*”). La fracción correspondiente a un 20 % de glicerol, en la que la banda analizada tenía un peso molecular aparente entre 66 y 97 kDa (según el perfil que presentaba una solución de marcadores de peso molecular que fueron analizados como control) fue depositada sobre una rejilla y analizada mediante microscopía electrónica (ver apartado 4.6 de Resultados).

#### 4.5. Análisis de la interacción de Xrcc4 con DNA

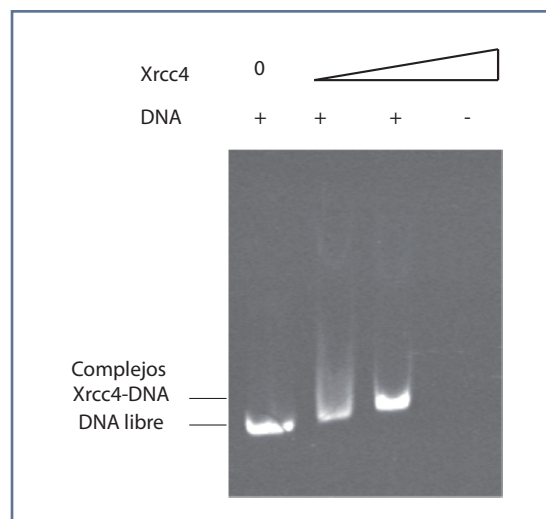
Con el objetivo de determinar si la proteína Xrcc4 purificada era capaz de unir DNA se realizaron ensayos de retardo en gel en condiciones nativas con un fragmento de



**Figura 4.5. Gel SDS-PAGE de las fracciones recogidas del gradiente de glicerol en el que se cargó la proteína Xrcc4.** La fracción seleccionada fue la que se utilizó para el posterior análisis mediante microscopía electrónica. Las bandas marcadas con “\*” corresponden a agregados de mayor peso molecular y contaminantes de la preparación. En azul, fracción del gradiente en el que eluía el marcador de peso molecular albúmina (66 kDa), y en rojo, fracción de elución del marcador fosforilasa b (97 kDa).

DNA de 390 pb de doble cadena. Las reacciones de unión se llevaron a cabo tal y como se describen en la sección 3.6 de Materiales y Métodos. Se incubó Xrcc4 con DNA en una relación molar 100:1 y 200:1, y a continuación, los complejos Xrcc4-DNA se resolvieron mediante electroforesis en condiciones nativas en un gel de poliácridamida. Como controles se analizaron las mismas cantidades de proteína y de DNA de manera independiente. Tras la electroforesis, el gel fue teñido con bromuro de etidio para detectar el retardo en el DNA (Figura 4.6). Cuando Xrcc4 y el DNA estaban presentes en la mezcla de reacción, se observó un retardo en el movimiento electroforético del DNA, en cuya banda también se observaba Xrcc4, determinándose que Xrcc4 purificada era capaz de unir DNA de doble cadena.

**Figura 4.6. Electroforesis en gel de poliácridamida en condiciones nativas para el análisis de la interacción de Xrcc4 con DNA.** Xrcc4, en concentraciones crecientes, se incubó con una molécula de DNA de 390 pb a una concentración fija y, a continuación, el complejo se resolvió mediante electroforesis en condiciones nativas. El gel fue teñido con bromuro de etidio para detectar el DNA donde se puede apreciar el retardo sufrido al interaccionar el DNA con la proteína.



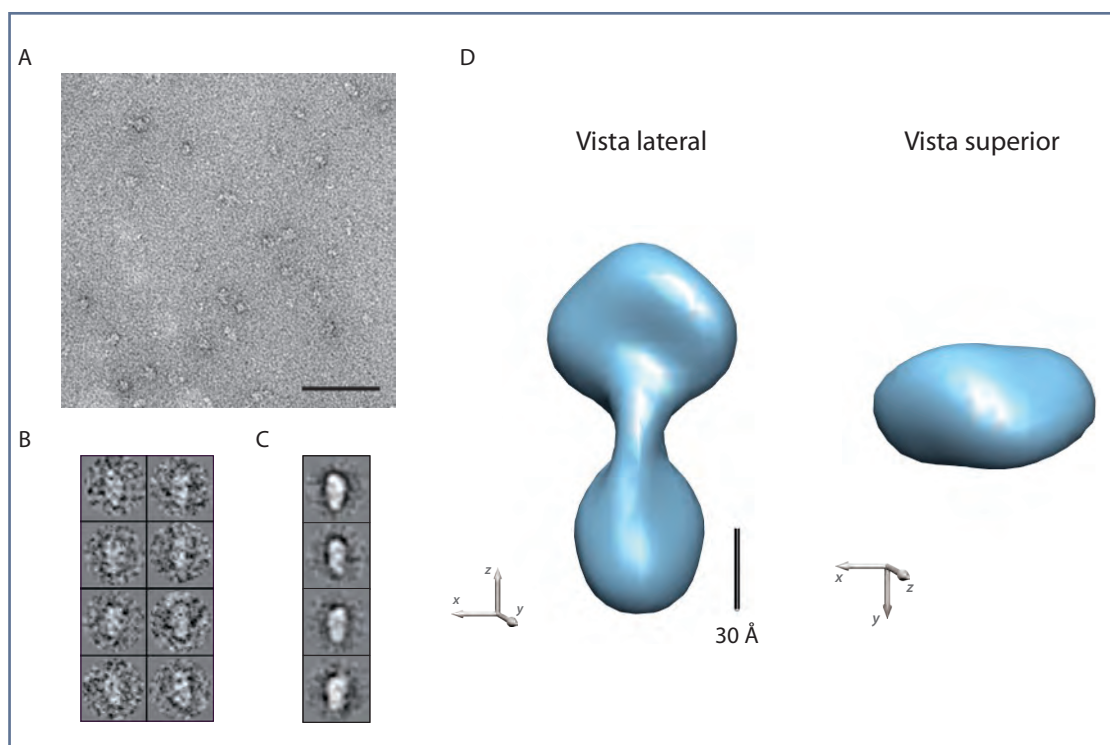


## 4.6. Análisis estructural de Xrcc4 de ratón

### 4.6.1. Reconstrucción tridimensional mediante microscopía electrónica

Como se ha descrito previamente en la sección de Introducción, han sido resueltas hasta la fecha tres estructuras atómicas de Xrcc4 de origen humano, pero todas ellas comprenden alrededor de 200 de los 336 aminoácidos que constituyen esta proteína (Ver Figuras 1.7 y 1.8 de Introducción) (Junop et al., 2000; Sibanda et al., 2001; Wu et al., 2009). Una cuarta estructura correspondiente a Lif1p (ortólogo de Xrcc4 en levaduras, del inglés *Ligase interacting factor 1*) expresada y purificada en *S.cerevisiae* ha sido resuelta formando un complejo con los dominios BRCT de la proteína Lig4p, ortólogo de DNA ligasa IV en levaduras (Dore et al., 2006). A pesar de los esfuerzos llevados a cabo por distintos grupos, hasta el momento no ha sido posible resolver la estructura de la proteína Xrcc4 completa. La estructura del dominio C-terminal de la proteína, que representa alrededor de un tercio de la molécula, permanece aún sin determinar, por lo que la organización tridimensional de la proteína completa se desconocía en el momento de realizar este trabajo.

Con el fin de determinar la arquitectura global del homodímero Xrcc4 de ratón completo se llevó a cabo un estudio mediante microscopía electrónica y análisis de partículas individuales. El homodímero de Xrcc4 tiene un peso molecular de sólo 76 kDa, por lo que se encuentra en el límite de lo que puede ser estudiado mediante esta técnica, y únicamente pudo ser analizado mediante tinción negativa. En las micrografías adquiridas se observaron moléculas alargadas (Figura 4.7, A), compatibles con las estructuras que habían sido determinadas previamente mediante cristalografía de rayos X. Se seleccionaron y extrajeron un total de 6916 partículas de las micrografías (Figura 4.7, B), las cuales fueron procesadas mediante métodos de alineamiento y promediado libres de referencia basados en correlación cruzada y máxima verosimilitud (ver apartado 3.8.3 de Materiales y Métodos) (Figura 4.7, C). Estas medias mostraron unas moléculas alargadas que correspondían con casi total probabilidad a vistas laterales de la proteína en distintas orientaciones rotacionales a lo largo de su eje longitudinal. Posteriormente, se llevó a cabo un análisis mediante procesos de refinamiento angular iterativo, generando una reconstrucción tridimensional. Para ello, se utilizaron exclusivamente las vistas más alargadas que podrían corresponder a vistas laterales de la molécula, descartándose las partículas más globulares. Además, se aplicó simetría C2 a lo largo del eje longitudinal de la molécula durante todo el procesamiento. El mapa obtenido de Xrcc4 resuelto a 37 Å de resolución (Figura 4.7, D) mostró que Xrcc4 es una molécula alargada de aproximadamente 130 Å de longitud constituida por un dominio globular en cada uno de sus dos extremos, siendo el de la parte superior de mayores dimensiones (70 Å de largo por 50 Å de ancho) que el situado en la parte inferior de la molécula (50 Å por 40 Å). El mapa, además, presentaba ciertas características estructurales que lo asemejaban al dominio N-terminal resuelto en la estructura atómica de

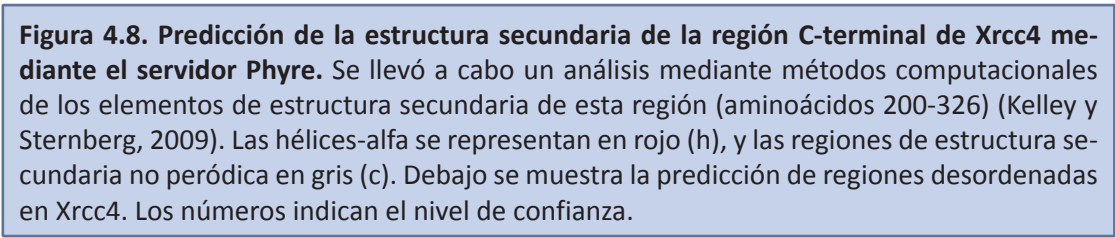


**Figura 4.7. Reconstrucción tridimensional de Xrcc4.** **A)** Campo correspondiente a una micrografía tomada de una rejilla de microscopía teñida negativamente con acetato de uranilo. La barra de escala corresponde a 100 nm. **B)** Galería de partículas individuales que se corresponden con dímeros de Xrcc4. **C)** Galería de imágenes medias obtenidas mediante métodos de clasificación libres de referencia. **D)** Vistas lateral y superior de la reconstrucción tridimensional de Xrcc4 obtenida.

Xrcc4, pero presentaba una masa adicional en uno de sus extremos. Las dos masas globulares situadas en los extremos de la molécula aparecían conectadas por una delgada densidad que podría corresponder al dominio *coiled-coil*.

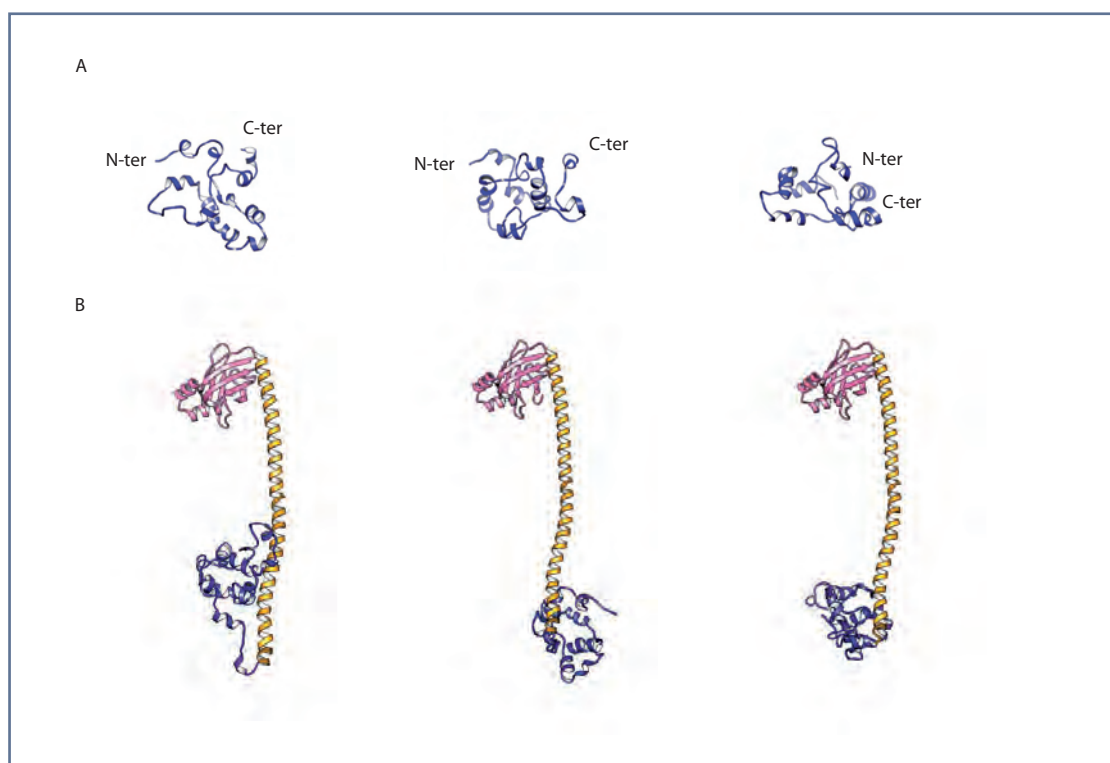
#### 4.6.2. Modelado del dominio C-terminal de Xrcc4

Debido a la ausencia de información a resolución atómica del dominio C-terminal de Xrcc4, se llevó a cabo un modelado de esta región. En primer lugar, se realizó una predicción de estructura secundaria de dicho dominio utilizando los servidores Robetta y Phyre (Figura 4.8). Ambos programas predijeron que el dominio está principalmente desordenado, pero presenta pequeñas regiones con un plegamiento en alfa-hélices. A continuación, con el objetivo de caracterizar y determinar la estructura del dominio C-terminal, se llevó a cabo una predicción de estructura terciaria de dicho dominio de Xrcc4 (Glu211-Asp326) utilizando el servidor Robetta. Se obtuvieron un conjunto de diez modelos estructurales posibles para Xrcc4, de los cuales tres de ellos son mostrados en la Figura 4.9, A. Las predicciones fueron realizadas *de novo* por Robetta, puesto que no existen homólogos estructurales para esta región. Los modelos obtenidos se caracterizaban por presentar una estructura de un solo dominio globular constituido



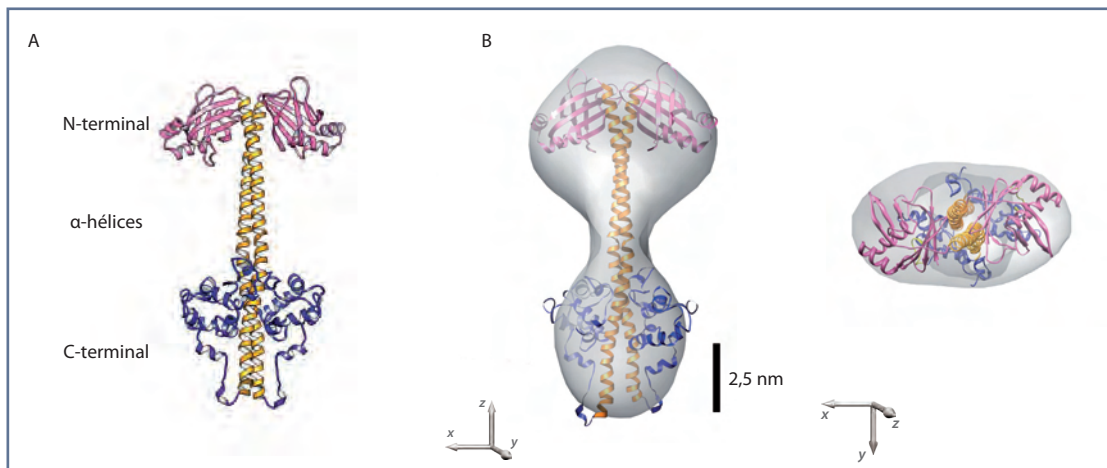
#### 4.6.3. Encaje de los modelos generados por Robetta en el mapa de densidad electrónica

56



**Figura 4.9. Modelado del dominio C-terminal de Xrcc4. A)** Resultados de la predicción de estructura tridimensional realizada con el servidor Robetta para el dominio C-terminal de Xrcc4 de ratón (se muestran los tres primeros resultados obtenidos). **B)** Resultados de la predicción de estructura tridimensional realizada con el servidor Robetta para la proteína completa de Xrcc4 (también se muestran los tres primeros resultados).

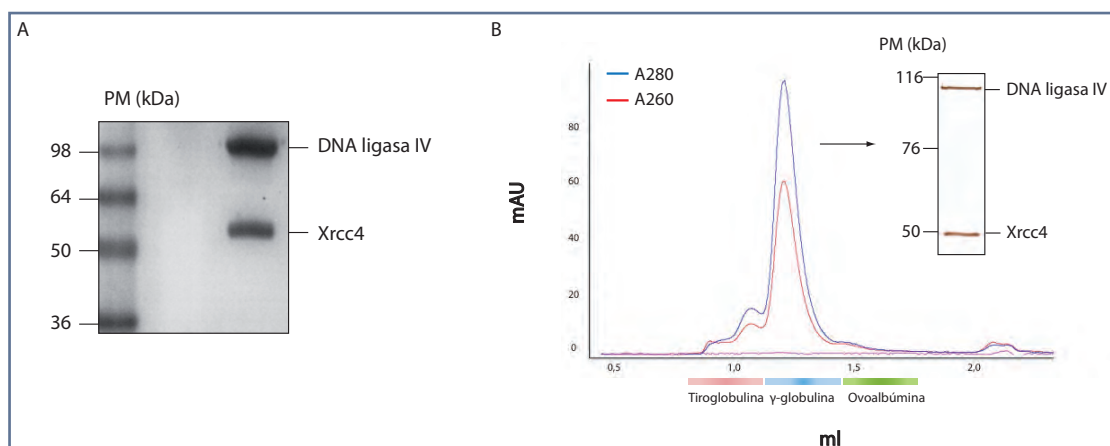
4.10, A). Estos dímeros fueron luego utilizados para realizar experimentos de ajuste en el mapa de densidad obtenido mediante el procesamiento de imágenes (Figura 4.10, B), utilizando par ello el programa informático ADP\_EM (Garzon et al., 2007). En este caso, las estructuras atómicas fueron filtradas a la resolución de la estructura obtenida mediante microscopía y se utilizó un método de correlación cruzada, en el cual se tiene en cuenta toda la densidad de ambas estructuras en el momento de hacer el ajuste. Las soluciones con mayor coeficiente de correlación revelaron que la región N-terminal de la estructura atómica de Xrcc4 se ajustaba de una manera adecuada en la región superior de la reconstrucción obtenida mediante microscopía, mientras que en la parte inferior podía ajustarse el dominio C-terminal del modelo atómico. Los modelos obtenidos con Robetta que mostraban el dominio C-terminal cercanos al dominio *coiled-coil* producían soluciones de encaje con un alto valor de correlación, mayor a 0,8. Por otro lado, los modelos en los que el dominio C-terminal se disponía más alejado de las hélices presentaban un menor coeficiente de correlación en el encaje.



**Figura 4.10. Encaje de la estructura obtenida mediante Robetta en el mapa de microscopía electrónica. A)** Modelo atómico obtenido mediante Robetta dimerizado, el cual obtuvo un mayor valor de correlación al ser encajado en el mapa de densidad generado mediante microscopía electrónica. **B)** Encaje de dicho modelo atómico en el mapa de microscopía.

#### 4.7. Purificación y caracterización del complejo Xrcc4-DNA ligasa IV humano

La purificación del complejo Xrcc4-DNA ligasa IV constituido por las formas completas de ambas proteínas se realizó a partir de células de insecto infectadas con baculovirus, siguiendo el protocolo descrito en el apartado 3.1.2 de Materiales y Métodos. Se purificó un complejo que fue analizado mediante electroforesis en condiciones desnaturalizantes, en la que se determinó que el complejo estaba puro, era homogéneo y estable (Figura 4.11, A). La composición del complejo purificado se confirmó mediante *western blot* utilizando anticuerpos específicos frente a Xrcc4 y DNA ligasa IV. En las primeras pruebas de microscopía llevadas a cabo observamos de nuevo fenómenos de agregación que impedían el análisis mediante esta técnica, al igual que en el caso de Xrcc4. A pesar de que la muestra se encontraba bioquímicamente pura, formaba agregados y se deformaba o aplastaba cuando se se preparaba para su observación al microscopio. Por este motivo, llevamos a cabo una purificación del complejo mediante cromatografía de filtración molecular en una columna Superdex 200 (Figura 4.11, B) con el objetivo de obtener una muestra más homogénea que mejorara su apariencia en el microscopio. El perfil de elución mostró que el complejo eluía en un pico con un peso molecular correspondiente a la molécula de  $\gamma$  globulina bobina (158 kDa), según los volúmenes de elución estandarizados de proteínas globulares de masa conocida. Las fracciones recogidas fueron analizadas mediante SDS-PAGE, y la fracción correspondiente al peso molecular del complejo Xrcc4-DNA ligasa IV (Figura 4.11, B, recuadro) se depositó sobre una rejilla y se analizó mediante microscopía electrónica.

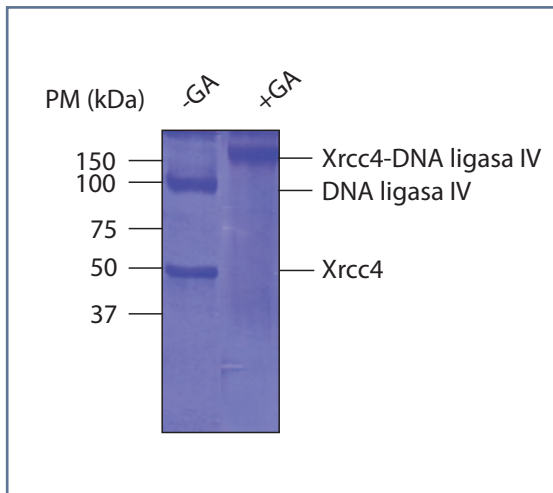


**Figura 4.11. Purificación del complejo Xrcc4-DNA ligasa IV. A)** Gel de poliacrilamida en condiciones desnaturalizantes realizado tras la purificación del complejo Xrcc4-DNA ligasa IV. PM: marcadores de peso molecular. **B)** Purificación del complejo mediante cromatografía de exclusión molecular previo a su análisis mediante microscopía electrónica. Perfil de elución del complejo y gel de poliacrilamida en condiciones desnaturalizantes correspondiente a la fracción seleccionada para su análisis mediante microscopía electrónica.

#### 4.8. Experimentos de entrecruzamiento del complejo Xrcc4-DNA ligasa IV mediante glutaraldehído

Con el fin de determinar el estado de oligomerización del complejo Xrcc4-DNA ligasa IV se llevó a cabo un ensayo de entrecruzamiento con glutaraldehído. Se incubó el complejo con un tampón que contenía diferentes concentraciones de glutaraldehído durante distintos tiempos (ver apartado 3.4 de Materiales y Métodos para más detalle). Las muestras se analizaron posteriormente mediante electroforesis en geles de poliacrilamida desnaturalizantes, donde se observó una banda mayoritaria que presentaba una migración electroforética correspondiente a una especie de aproximadamente 180 kDa. Este peso molecular resultó ser compatible con un complejo Xrcc4-DNA ligasa IV en relación molar 2:1, correspondiente a una molécula de DNA ligasa IV por cada dímero de Xrcc4 (Figura 4.12). Como control se cargó la misma cantidad de proteína sin glutaraldehído, detectándose las dos bandas correspondientes a Xrcc4 y la DNA ligasa. Además, se detectaron otros componentes de mayor peso molecular, probablemente correspondientes a agregados que fueron observados de forma repetida en los estudios de microscopía realizados. La banda que probablemente correspondía al complejo Xrcc4-DNA ligasa IV fue analizada mediante espectrometría de masas, que determinó que ambas proteínas estaban presentes.

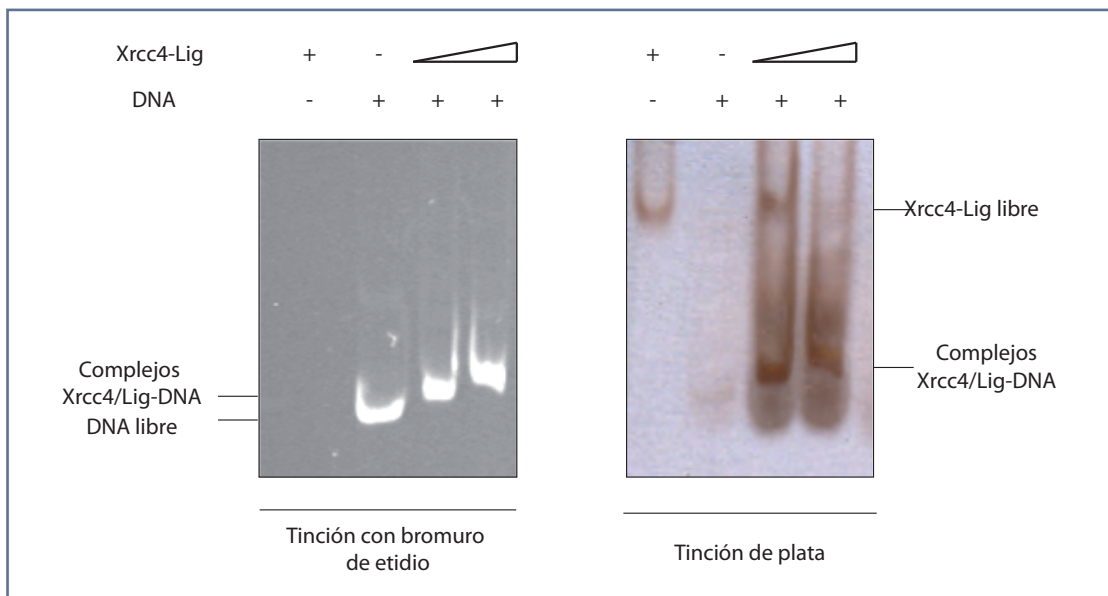




**Figura 4.12. Experimentos de entrecruzamiento del complejo Xrcc4-DNA ligasa IV.** Gel de poliacrilamida en condiciones desnaturalizantes en el que se resolvieron los complejos después del ensayo de entrecruzamiento con glutaraldehído. El peso molecular de los marcadores se indica en la parte izquierda en kDa. En la parte derecha se indica la posición de la principal especie detectada. Como control, se cargó una muestra del mismo complejo que no había sido tratada previamente con glutaraldehído.

#### 4.9. Análisis de la interacción del complejo Xrcc4-DNA ligasa IV con DNA

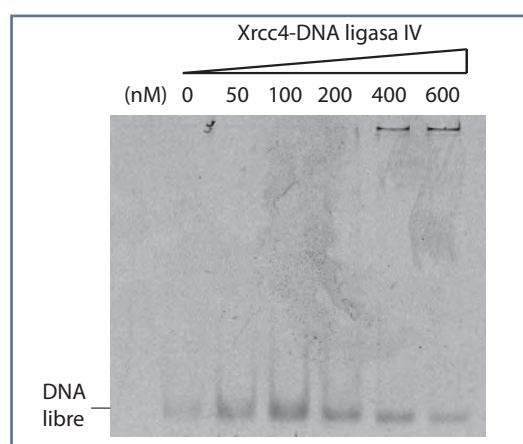
Con el fin de determinar si el complejo Xrcc4-DNA ligasa IV interaccionaba con DNA se llevaron a cabo ensayos de retardo en gel en condiciones nativas con dos tipos distintos de DNA. Por un lado, se incubó el complejo con el fragmento de DNA de 390 pb de doble cadena descrito en Materiales y Métodos en una relación molar 30:1 y 60:1 con el objetivo de que la mayor parte del DNA se uniera a proteína, y a continuación, los complejos Xrcc4-DNA se resolvieron mediante electroforesis en condiciones nativas en geles de poliacrilamida. Como controles se analizaron las mismas cantidades de proteína y de DNA de manera independiente. El gel fue teñido con bromuro de etidio con el fin de detectar el DNA (Figura 4.13) y en segundo lugar con una solución de pla-



**Figura 4.13. Ensayo de interacción entre Xrcc4-DNA ligasa IV y DNA de 390 pb.** A) Gel nativo teñido con bromuro de etidio, en el cual se observa retardo en el DNA. B) El mismo gel nativo fue teñido posteriormente con plata, donde se observó que la movilidad electrofórica era diferente en ausencia y presencia de DNA.

ta, con el fin de determinar la diferencia de movilidad electroforética en el complejo. Se observó un retardo en el movimiento electroforético del DNA, en cuya banda también aparecía el complejo. Además, al igual que en el caso de Xrcc4, se observó que los complejos con DNA poseían una mayor movilidad electroforética que la proteína libre.

Por otro lado, también se llevó a cabo la unión del complejo a una molécula de DNA de 35 pb, analizando una concentración de DNA constante de 25 nM con concentraciones crecientes del complejo, desde 50 nM a 600 nM. En este caso, observamos retardo en el complejo cuando la relación molar DNA:complejo era entre 1:8 y 1:16, pero en ningún caso se observó el retardo completo del DNA (Figura 4.14).

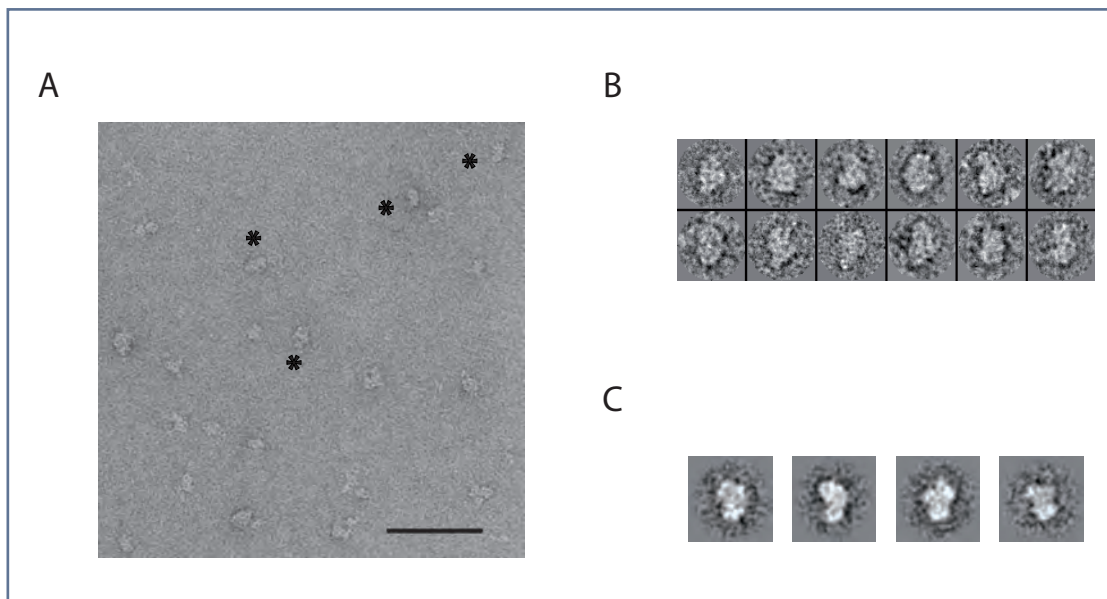


**Figura 4.14. Ensayo de interacción entre el complejo Xrcc4-DNA ligasa IV y una molécula de DNA de 35 pb.** Gel nativo en el que se cargaron distintas concentraciones del complejo (0-600 nM), con una cantidad constante de DNA (25 nM). El gel fue visualizado en un Phosphorimager, y en él se puede observar el retardo en el DNA cuando las concentraciones de complejo Xrcc4-DNA ligasa IV se encuentran entre 8:1 y 16:1 con respecto al DNA.

#### 4.10. Análisis estructural del complejo Xrcc4-DNA ligasa IV mediante microscopía electrónica

Una vez determinada la homogeneidad y la composición de la preparación, se procedió a la visualización de la muestra en el microscopio electrónico. Las partículas del complejo Xrcc4-DNA ligasa IV fueron observadas después de teñirlas negativamente, siguiendo el protocolo descrito en el apartado 3.8.1 de Materiales y Métodos. A pesar de haber realizado una cromatografía de filtración molecular para separar el complejo de los agregados, en las micrografías adquiridas seguía observándose una gran variabilidad en cuanto a tamaños y formas de las moléculas (Figura 4.15, A). Aún así, se detectaron partículas individuales del tamaño esperado para el complejo distribuidas sobre la rejilla. Se seleccionaron y extrajeron un total de 7848 imágenes individuales de las micrografías, algunas de las cuales se muestran en la Figura 4.15, B, y se procesaron mediante métodos de clasificación y alineamiento libres de patrón descritos en el apartado 3.8.3 de Materiales y Métodos (Figura 4.15, C). Cada una de estas imágenes medias se corresponde con una imagen de proyección de la partícula bajo estudio, la cual ha interactuado con el soporte de microscopía electrónica en una orientación determinada. Estas imágenes mostraron que las moléculas presentaban nuevamente una forma alargada al igual que Xrcc4, pero se observó que poseían masa adicional en uno de sus extremos, lo cual no había sido observado previamente en las moléculas de



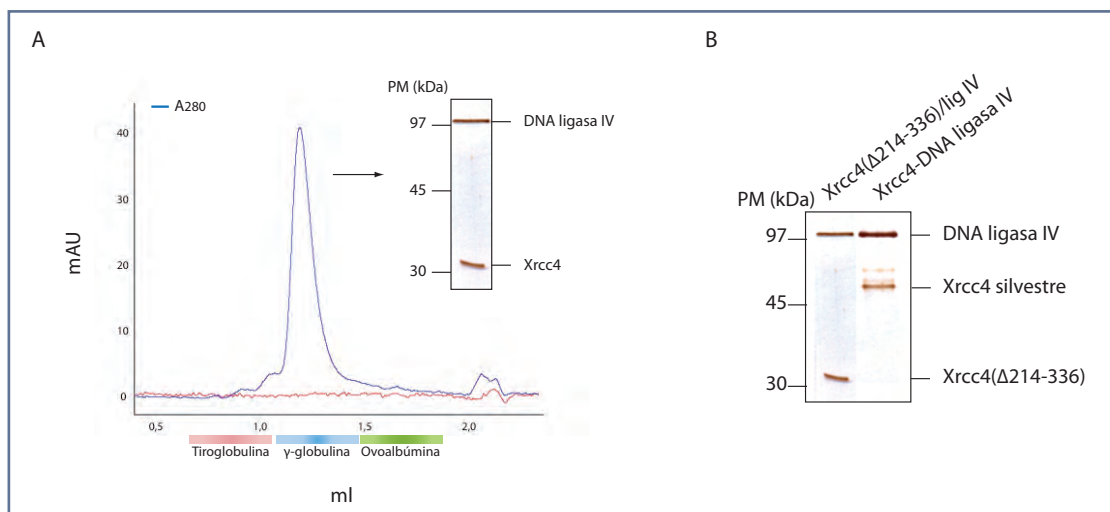


**Figura 4.15. Microscopía electrónica de Xrcc4-DNA ligasa IV.** **A)** Campo de una micrografía electrónica de una muestra del complejo contrastado negativamente. La barra de escala indica 100 nm. Los asteriscos indican partículas individuales correspondientes a complejos Xrcc4-DNA ligasa IV. **B)** Galería de imágenes correspondientes a moléculas individuales representativas extraídas de las micrografías. **C)** Clases medias obtenidas mediante algoritmos libres de referencia.

Xrcc4 cuando la proteína DNA ligasa IV no se encontraba presente en la preparación.

#### 4.11. Localización del extremo C-terminal de Xrcc4 en el complejo Xrcc4-DNA ligasa IV

Con el fin de localizar la posición del dominio C-terminal de Xrcc4 en el complejo Xrcc4-DNA ligasa IV se purificaron y observaron en el microscopio electrónico complejos que carecían de dicha región. Para llevarlo a cabo, se purificó un complejo constituido por la proteína DNA ligasa IV completa y una versión truncada de Xrcc4 ( $\Delta 214-336$ ) (ver apartado 3.1.3 de Materiales y Métodos). Mediante cromatografía de exclusión molecular se determinó que la movilidad del complejo era similar a la del marcador y globulina bovina, el cual tiene un peso molecular de 158 kDa (Figura 4.16, A). Las fracciones recogidas se analizaron mediante SDS-PAGE, donde se pudo observar la pureza de la preparación, así como la diferencia en la movilidad electroforética para Xrcc4 truncada comparada con aquella que formaba parte del complejo silvestre (Figura 4.16, B). A continuación, la fracción correspondiente al pico en el que eluía el complejo se analizó mediante microscopía electrónica. Se seleccionaron y extrajeron un total de 3428 partículas (Figura 4.17, A), las cuales fueron promediadas generando medias libres de referencia del complejo truncado (Figura 4.17, B). Éstas mostraron una molécula significativamente más pequeña y globular comparada con el complejo silvestre, lo cual sugería que la pérdida del dominio C-terminal de Xrcc4 modificaba

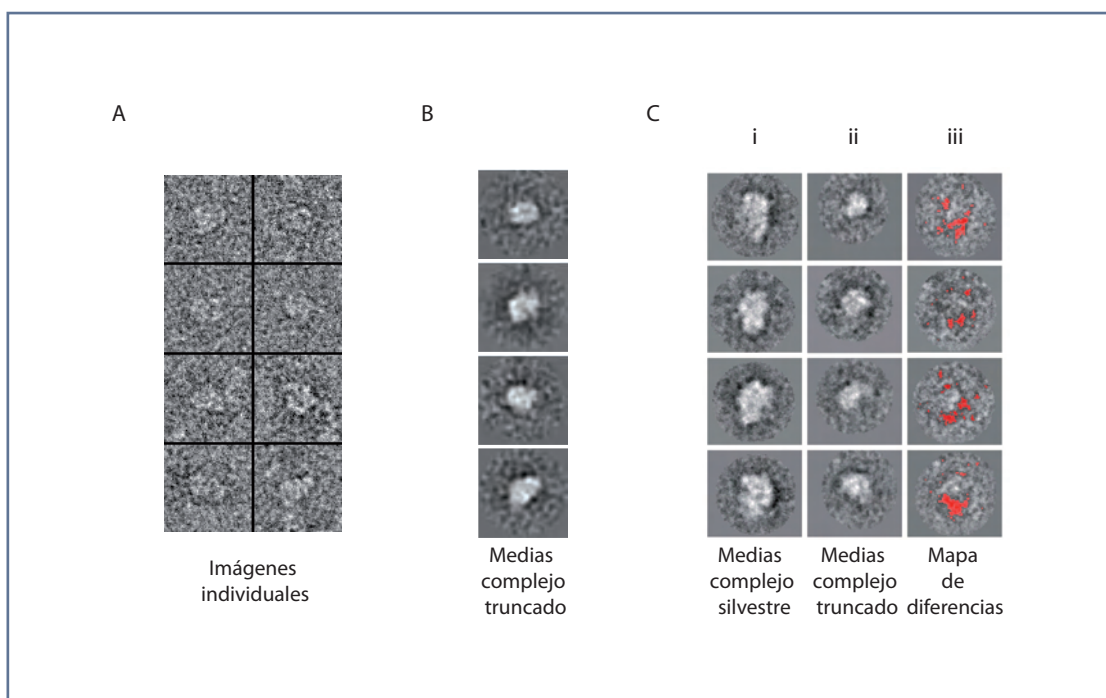


**Figura 4.16. Purificación del complejo Xrcc4( $\Delta$ 214-336)/DNA ligasa IV. A)** Cromatografía de exclusión molecular en la que se analizó el estado de agregación del complejo. Como marcadores de peso molecular se utilizó una solución de proteínas de movilidad conocida, entre ellas tiroglobulina (670 kDa),  $\gamma$ -globulina (158 kDa) y ovoalbúmina (45 kDa). En el recuadro se muestra el gel SDS-PAGE en el que se analizó la fracción de la cromatografía correspondiente al complejo. **B)** Gel SDS-PAGE en el que se comparó el complejo truncado con el complejo silvestre. Se puede observar la diferencia en la movilidad electroforética entre el complejo silvestre y el truncado.

la forma alargada de la molécula al visualizarla en el microscopio electrónico. Esta reducción en la longitud del complejo después de eliminar tan sólo un pequeño dominio únicamente podía ser compatible con la localización del extremo C-terminal de Xrcc4 en el extremo inferior más estrecho de las imágenes. A continuación se llevó a cabo un alineamiento de las imágenes medias obtenidas del complejo truncado con medias generadas previamente con los datos del complejo silvestre (Figura 4.17, C i, ii), cuyo resultado mostró que la mayoría de las medias eran alineadas con la región más ancha de dichas medias. Posteriormente, se calculó el mapa de diferencias entre las imágenes medias alineadas de las clases correspondientes a ambos conjuntos de datos (Figura 4.17, Ciii), el cual localizó la región truncada de Xrcc4 en el extremo más estrecho de las imágenes medias bidimensionales del complejo Xrcc4-Dna ligasa IV silvestre. Por tanto, la región N-terminal de Xrcc4 y la ligasa probablemente se localizaban en la parte más voluminosa de las imágenes medias.

A continuación, se intentó llevar a cabo la reconstrucción tridimensional del complejo truncado, pero debido a esta reducción en tamaño así como su aspecto globular no fue posible resolver su estructura. Sin embargo, los volúmenes preliminares que se obtuvieron durante el procesamiento sólo se podían interpretar asignando el mapa de densidad obtenido a la parte más voluminosa y globular del complejo Xrcc4-DNA ligasa IV silvestre, lo cual es compatible con los resultados obtenidos con anterioridad a partir de las imágenes bidimensionales.

Por tanto, con este ensayo se ha localizado la posición del extremo C-terminal de



**Figura 4.17. Microscopía electrónica del complejo Xrcc4( $\Delta$ 214-336)/DNA ligasa IV y mapeo del extremo C-terminal de Xrcc4 en el complejo. A)** Galería de partículas individuales extraídas de las micrografías. **B)** Imágenes medias del complejo truncado obtenidas tras el procesamiento mediante algoritmos libres de patrón. **C)** Correlación entre las medias libres de referencia del complejo truncado con medias del complejo silvestre. Cada línea muestra: **i)** medias 2D del complejo silvestre; **ii)** medias 2D del complejo truncado después de ser alineadas frente a las medias mostradas en i; **iii)** mapa de diferencia entre las imágenes alineadas del complejo Xrcc4-DNA ligasa IV truncado y el complejo silvestre (con las diferencias significativas mostradas en rojo).

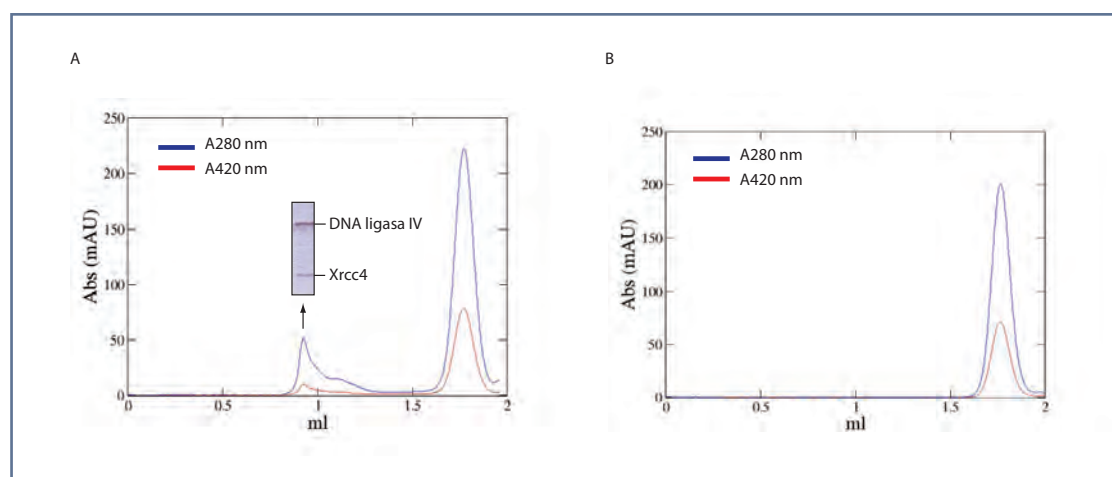
Xrcc4 en las imágenes del complejo Xrcc4-DNA ligasa IV.

#### 4.12. Localización del extremo N-terminal de la DNA ligasa IV en el complejo Xrcc4-DNA ligasa IV mediante marcaje con moléculas de oro

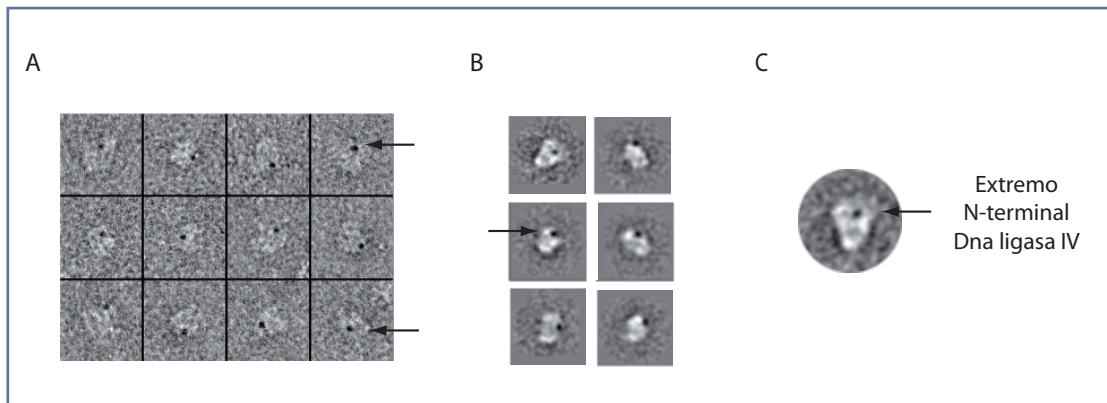
Con el fin de determinar la localización del núcleo catalítico de la ligasa en las imágenes del complejo, llevamos a cabo un experimento de marcaje con oro en el complejo Xrcc4-DNA ligasa IV silvestre. Se incubó el complejo con un compuesto que contenía una partícula de oro de 1,8 nm de diámetro unida a un átomo de níquel (Ni-NTA Nanogold, ver sección 3.11 de Materiales y Métodos). Este compuesto se une con una alta afinidad a varios residuos de histidina adyacentes, y la ligasa del complejo presentaba una etiqueta accesible de seis histidinas en su dominio N-terminal. El complejo se incubó con un exceso de Nanogold y los complejos resultantes se purificaron mediante cromatografía de exclusión molecular. El cromatograma reveló que la proteína y el oro coelúan en un único pico identificado mediante absorbancia a 280 nm y 420 nm (Figura 4.18, A), siendo la absorbancia a 420 nm específica para detectar la presencia de partículas de oro. Como controles se inyectaron en la columna las mismas cantidades

de Nanogold y del complejo Xrcc4-DNA ligasa IV de manera independiente, y se analizaron y compararon los cromatogramas con los obtenidos en el primer caso (Figura 4.18, B). Así mismo, se observó que la coelución de oro y proteína no ocurría en presencia de imidazol, confirmándose así que la interacción oro-proteína era específica de la etiqueta de histidinas presente en el dominio N-terminal de la ligasa. La proporción relativa de unión del oro a la proteína, calculada mediante los valores de absorbancia obtenidos, se estimó próxima al 65%. Se llevó a cabo a continuación una electroforesis en geles de poliacrilamida en condiciones desnaturalizantes (SDS-PAGE) que confirmó la existencia del complejo Xrcc4-DNA ligasa IV en la fracción analizada. A continuación se analizó la fracción que contenía el complejo unido a Nanogold mediante microscopía electrónica (Figura 4.19), donde se observó que un gran número de las imágenes correspondían a complejos proteína-oro, identificados por la presencia de un círculo de alta densidad a los electrones, y tenían un tamaño, forma y apariencia compatible con una partícula de oro de 1,8 nm (indicado con flechas en la Figura 4.19, A y B). Se seleccionaron y extrajeron de las micrografías un total de 1500 imágenes en las que claramente se detectó la partícula de oro y se generaron imágenes medias libres de patrón (Figura 4.19, B), las cuales indicaron que el motivo de alta densidad a los electrones correspondiente a la partícula de oro estaba localizado en las proximidades de uno de los extremos de la molécula, concretamente, en el extremo más ancho, donde mediante el análisis estructural llevado a cabo previamente en 2D se había conjeturado que se podía localizar la región N-terminal de Xrcc4 y parte de la DNA ligasa IV.

Por tanto, con este experimento de marcaje, se corroboró que en la región más extensa de la molécula se encontraba tanto la región N-terminal de Xrcc4 como al menos el núcleo catalítico de la ligasa, puesto que es en esa zona donde se encontraba el extremo N-terminal de la proteína (Figura 4.19, C).



**Figura 4.18. Purificación de complejos Xrcc4-DNA ligasa IV con Nanogold mediante cromatografías de exclusión molecular. A)** Cromatografía de exclusión molecular tras la incubación del complejo con Nanogold. **B)** Cromatograma obtenido tras inyectar una solución de Nanogold a la misma concentración que en A.

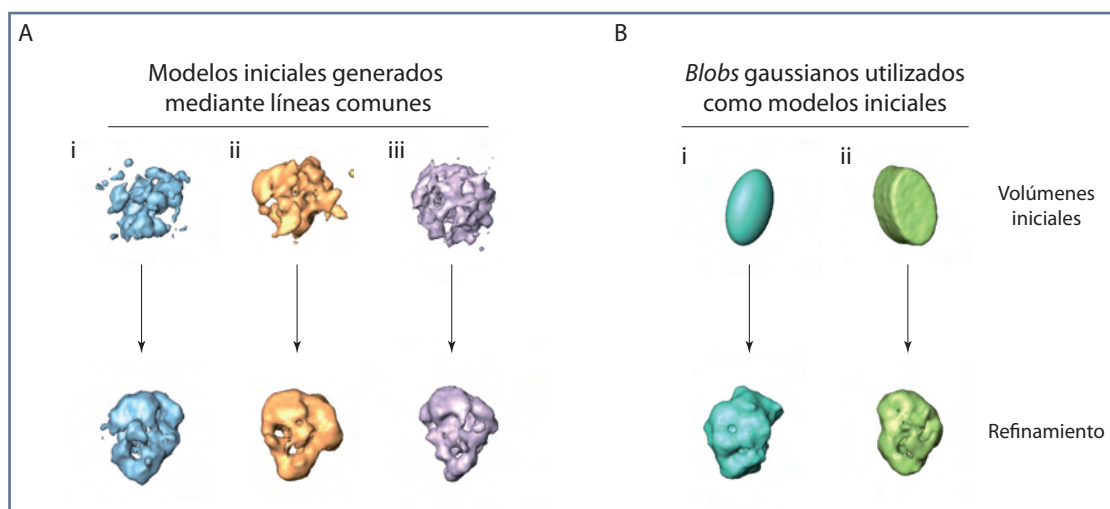


**Figura 4.19. Localización del núcleo catalítico de la proteína DNA ligasa IV dentro del complejo silvestre.** **A)** Partículas individuales extraídas de las micrografías en las que se observó claramente la molécula de oro. **B)** Imágenes medias libres de patrón generadas con las moléculas seleccionadas. **C)** Representación de la localización del núcleo catalítico de la ligasa dentro del complejo, utilizando una de las medias bidimensionales representativas.

#### 4.13. Reconstrucción tridimensional del complejo Xrcc4-DNA ligasa IV

Las partículas seleccionadas procedentes de la muestra Xrcc4-DNA ligasa IV silvestre fueron analizadas y procesadas utilizando métodos de refinamiento angular. Como se ha descrito previamente, se utilizaron diferentes estrategias de procesamiento para la generación del volumen inicial, todos ellos carentes de simetría, con los que se obtuvieron soluciones similares (Figura 4.20). Por un lado, las imágenes medias obtenidas mediante métodos de clasificación libres de patrón fueron utilizadas para generar un modelo tridimensional inicial mediante líneas comunes (Figura 4.20, A). Por otro lado, se generaron blobs gaussianos con las dimensiones del complejo y sin apenas detalles estructurales (Figura 4.20, B). Los modelos iniciales se refinaron de manera iterativa durante varios ciclos con EMAN y en un momento determinado, cuando se observó que los resultados eran similares, uno de los volúmenes procedente del refinamiento fue seleccionado y procesado con Xmipp.

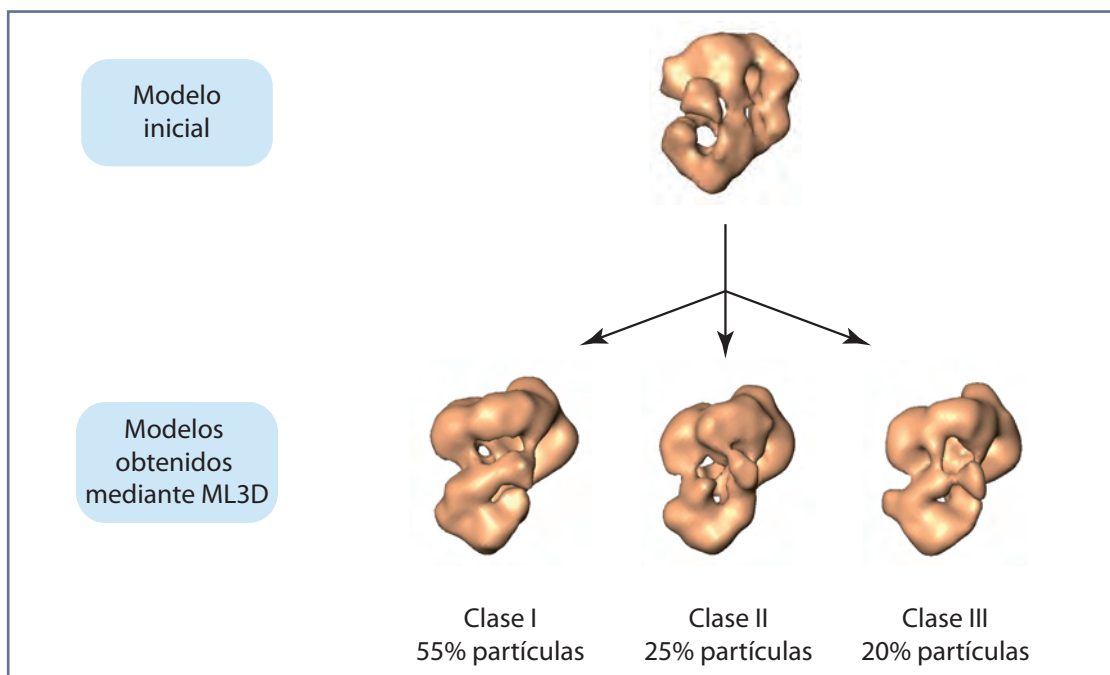
A continuación, se llevó a cabo un refinamiento con ML3D de Xmipp, en el que se calcularon tres reconstrucciones tridimensionales (Figura 4.21) en un intento de seleccionar un subconjunto de datos consistente, ya que se había observado que los datos poseían cierto grado de heterogeneidad estructural. Tras el refinamiento, un 55% de las partículas encajaron con el primer modelo obtenido, un 25% con el segundo y por último, un 20% de las mismas con el tercero. Sin embargo, los volúmenes generados resultaron ser similares entre sí, aunque ligeramente diferentes con respecto al mapa inicial. El volumen con un mayor número de partículas (conformación 1) fue seleccionado y refinado posteriormente, aunque este volumen no se estabilizó. Por este motivo, la reconstrucción obtenida no se puede considerar como una estructura definitiva del complejo. Probablemente, la estructura resuelta representa una media de distintas conformaciones del complejo que no ha sido posible resolver con las técnicas que he-



**Figura 4.20. Modelos iniciales empleados en el proceso de reconstrucción.** Se generaron distintos modelos iniciales, entre los que se incluían **A)** modelos generados mediante líneas comunes, **B)** blobs gaussianos. En todos los casos, las estructuras convergían hacia estructuras con características similares a baja resolución.

mos aplicado durante el procesamiento de imagen.

Sin embargo, el mapa obtenido era consistente con las medias libres de patrón generadas con los datos, lo cual sugiere que ciertas características estructurales del modelo podrían ser fidedignas. La estructura reveló una organización asimétrica, con unas



**Figura 4.21. Mapas obtenidos tras el procesamiento con ML3D de Xmipp.** Se introdujeron tres referencias iniciales en el procesamiento y las partículas se distribuyeron entre ellos del modo que se indica en la figura.

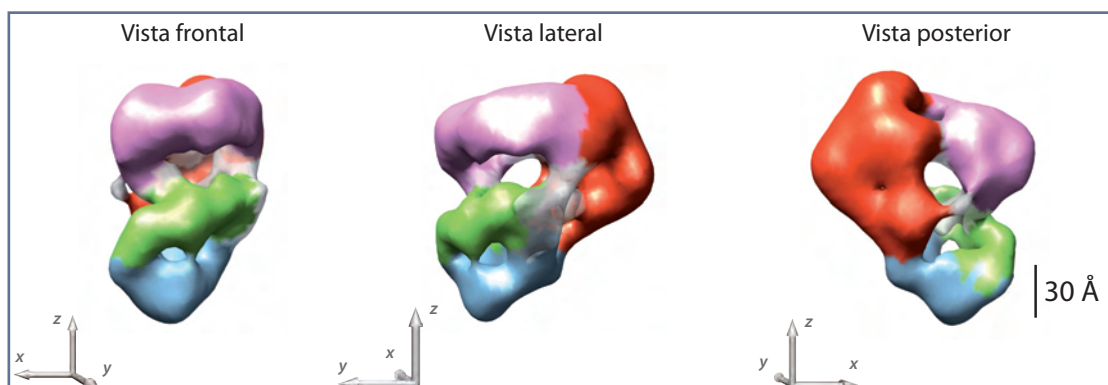


dimensiones aproximadas de 140 Å de alto por 110 Å de largo (máxima longitud) y 80 Å de ancho (Figura 4.22), y una resolución de 31 Å. En el mapa de densidad obtenido se distinguieron tres zonas diferenciadas. Por un lado, una región alargada constituida por dos masas en los extremos (destacados en colores rosa y azul) conectadas por una región más delgada (Figura 4.22, vista frontal). Se observó que tanto las dimensiones como la forma de esta región del modelo poseían características similares a la estructura atómica de Xrcc4. Por otro lado, se pudo apreciar una masa de mayor tamaño que atravesaba las anteriores (color verde en Figura 4.22, vista frontal). Se observó además una tercera masa de gran tamaño (de color rojo en la figura), la cual se encontraba adyacente a la masa de color rosa mencionada con anterioridad, y que era claramente visible desde otras orientaciones del mapa (Figura 4.22, vistas lateral y posterior).

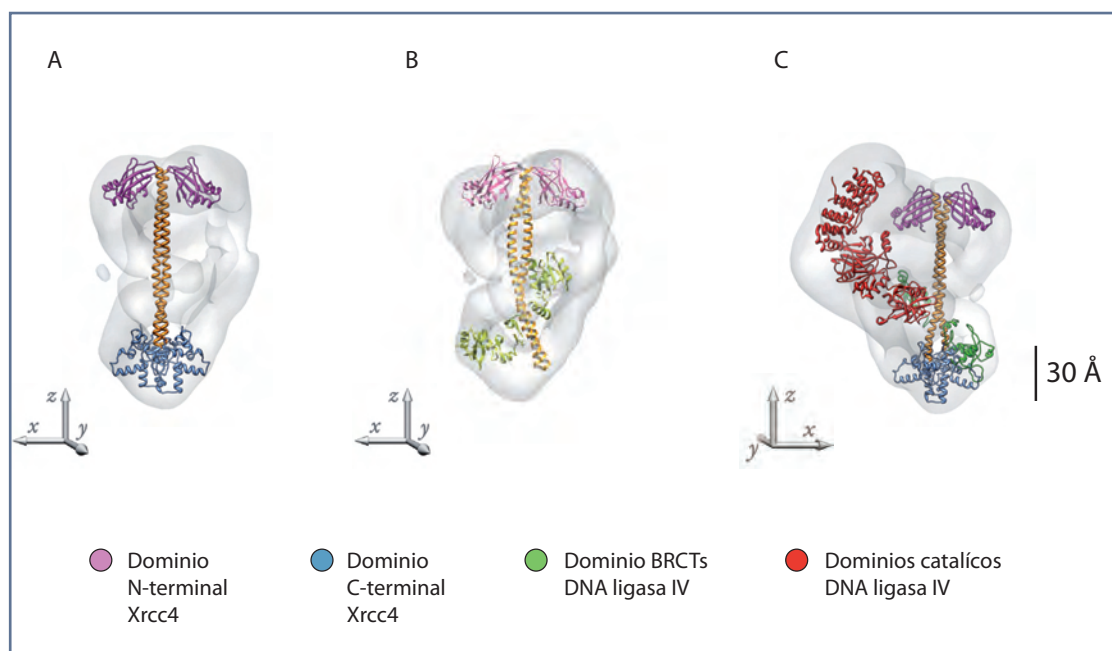
A continuación, llevamos a cabo el ajuste de diversas estructuras atómicas resueltas para determinados dominios del complejo o de otras ligasas homólogas en el volumen obtenido mediante microscopía electrónica (Figura 4.23). Al igual que con Xrcc4, también se generaron modelos atómicos de Xrcc4 humana completa utilizando el servidor Robetta, los cuales resultaron ser muy similares a los obtenidos con Xrcc4 de ratón, y que fueron utilizados también en el encaje en el mapa obtenido mediante microscopía electrónica (Figura 4.23, A).

Todas estas estructuras encajaban adecuadamente en el volumen y venían a confirmar los datos obtenidos previamente con los experimentos de localización en dos dimensiones de la región N-terminal de la ligasa y la región C-terminal de Xrcc4.

Por un lado, se llevó a cabo el ajuste del núcleo catalítico de una DNA ligasa dependiente de ATP del organismo *Sulfolobus solfataricus* homóloga a la proteína DNA ligasa IV (fichero PDB: 2HIV) (Figura 4.23, C) (Pascal et al., 2006). Esta estructura encajaba en la zona que se había asignado a la ligasa mediante el ensayo de localización con Nano-gold, mientras que el modelo de Xrcc4 generado con Robetta sólo podía ser encajado



**Figura 4.22. Reconstrucción tridimensional del complejo Xrcc4-DNA ligasa IV.** Este volumen fue utilizado para el encaje posterior de las estructuras atómicas. Se muestran distintas orientaciones de la molécula (frontal, lateral y posterior relacionadas por un ángulo de 90° entre ellas).



**Figura 4.23. Encaje de las estructuras atómicas disponibles para el complejo Xrcc4-DNA ligasa IV.** **A)** Encaje de la estructura modelada mediante Robetta para Xrcc4 humana completa. **B)** Encaje de la estructura resuelta mediante cristalografía de rayos X de Xrcc4 humana unida a los dominios BRCT de la ligasa (Wu et al., 2009) (fichero PDB 3I16). **C)** Encaje de la estructura obtenida mediante Robetta de Xrcc4 (colores rosa, naranja y azul), los dominios BRCT de DNA ligasa IV humana, extraídos del PDB 3I16 (en color verde) y el núcleo catalítico de la DNA ligasa de *Sulfolobus solfataricus* (en color rojo) (Pascal et al., 2006) (fichero PDB 2HIV).

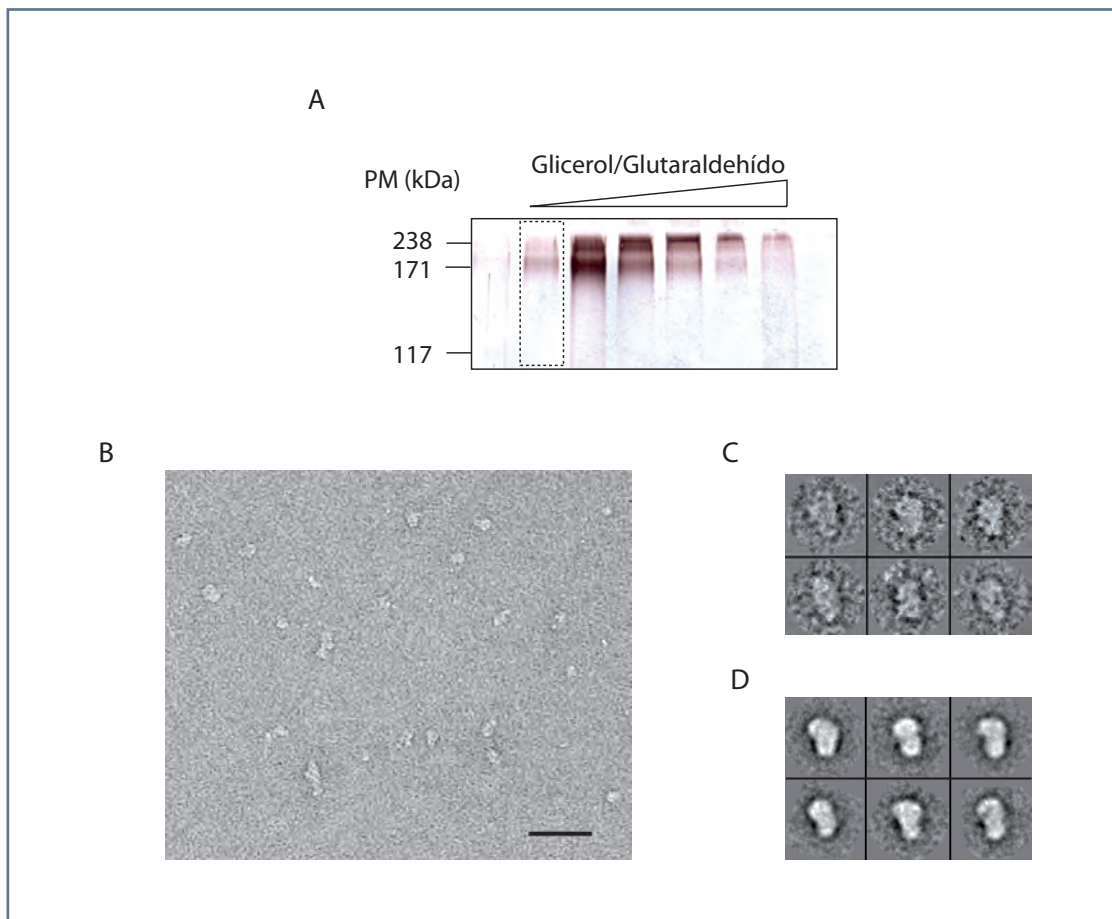
dentro del volumen en el eje longitudinal más largo de la estructura (Figura 4.23, A). Por último, los dominios BRCT pertenecientes a la proteína DNA ligasa IV de origen humano, extraídos del fichero PDB 3I16 (Wu et al., 2009), encajaban perfectamente en la región bilobulada central de la molécula, y además esta disposición concordaba con la estructura atómica resuelta de Xrcc4 en interacción con los dominios BRCT de la DNA ligasa IV, interacción que se establece a través de la región *coiled-coil* de Xrcc4 (Figura 4.23, B y C). La estructura obtenida era consistente, por tanto, con el mapeo realizado previamente en dos dimensiones, así como con las estructuras atómicas disponibles. Sin embargo, el hecho de que la estructura no se estabilizase durante el refinamiento sugería la presencia de heterogeneidad en los datos.

Para intentar aumentar la homogeneidad de la muestra y obtener así una reconstrucción con una mejor calidad de los datos, decidimos abordar estrategias de purificación de complejos Xrcc4-DNA ligasa IV estabilizados con glutaraldehído siguiendo el método GraFix.



#### 4.14. Purificación del complejo Xrcc4-DNA ligasa IV estabilizado mediante glutaraldehído (GraFix)

A continuación se llevó a cabo la purificación del complejo Xrcc4-DNA ligasa IV incubado con glutaraldehído siguiendo el método GraFix, con la intención de purificar un complejo más estable. Se sometió a la muestra a un proceso de centrifugación en un gradiente de glicerol en presencia de un gradiente creciente de glutaraldehído, que produce la fijación de las diferentes especies presentes a lo largo del gradiente. Las fracciones recogidas después del proceso de centrifugación se analizaron mediante electroforesis en geles de poliacrilamida en condiciones desnaturalizantes (Figura 4.24, A). Se observó una transición a lo largo de las fracciones desde complejos de unos 180 kDa hasta especies de mayor tamaño, probablemente correspondientes a agregados inducidos por el agente de entrecruzamiento. Se seleccionó una fracción enriquecida en complejos con un peso molecular aproximado de 180 kDa, compatible



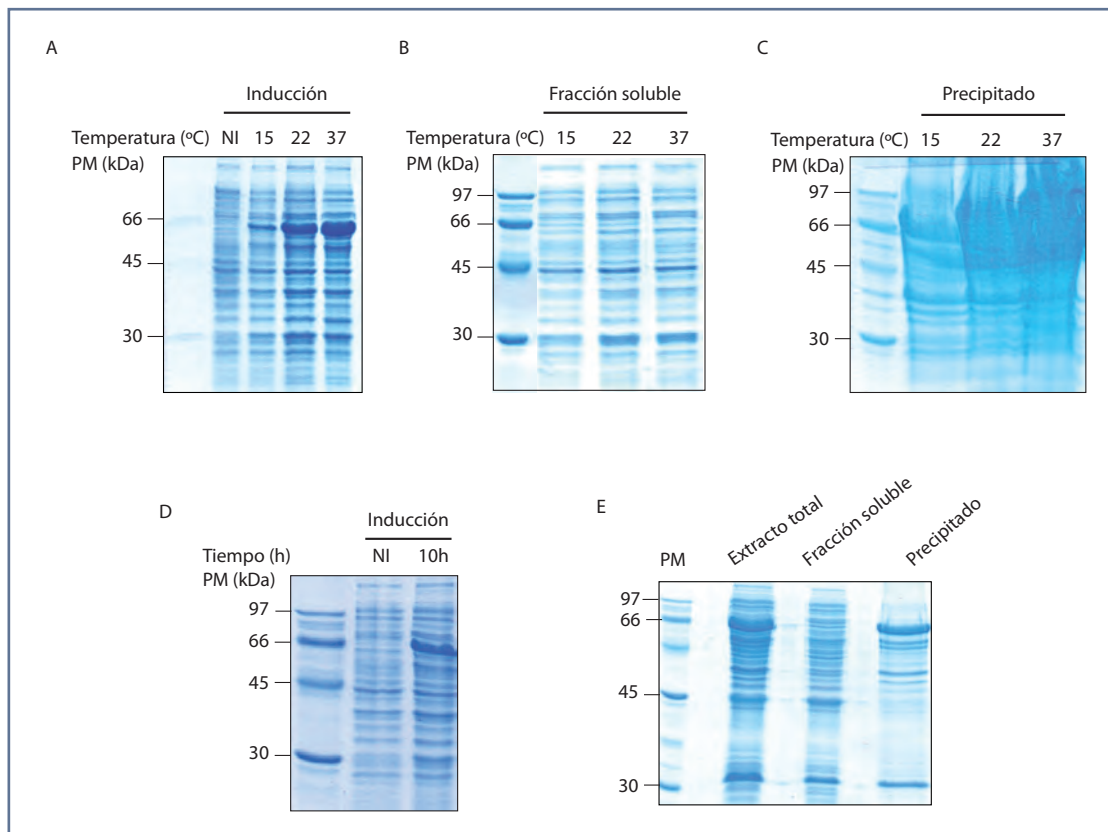
**Figura 4.24. Ensayo de GraFix del complejo Xrcc4-DNA ligasa IV silvestre. A)** Gel SDS-PAGE en el que se cargaron las fracciones recogidas tras GraFix. La fracción seleccionada, y que presentaba un peso molecular de aproximadamente 180 kDa fue analizada mediante microscopía electrónica. **B)** Campo de una micrografía en el que se muestran moléculas individuales. La barra de escala indica 50 nm. **C)** Galería de imágenes individuales extraídas de las micrografías. **D)** Clases medias libres de patrón.

con una molécula de ligasa unida a un dímero de Xrcc4. Dicha fracción se observó al microscopio electrónico, donde se tomaron micrografías (Figura 4.24, B) y se extrajeron 3519 partículas, algunas de las cuales pueden observarse en la Figura 4.24, C. Posteriormente, los datos se analizaron mediante procesamiento de imágenes, llevándose a cabo un análisis bidimensional libre de referencia para alinear las partículas de la misma manera realizada anteriormente (Figura 4.24, D). Estas medias mostraron que la molécula presentaba las mismas dimensiones y una morfología general similar a la descrita previamente para el caso de las moléculas del complejo sin fijación, sin embargo las partículas se habían teñido mejor que en el caso anterior y presentaban una mayor calidad. Se observó que las medias bidimensionales típicas del complejo Xrcc4-DNA ligasa IV consistía en una molécula longitudinal con un extremo más estrecho en la parte inferior y un extremo significativamente más ancho en el otro extremo, como ya se había determinado inicialmente con los datos obtenidos sin utilizar GraFix. Con estos nuevos datos se intentó llevar a cabo una reconstrucción tridimensional, sin embargo, el mapa obtenido no reveló nuevos detalles internos de la molécula, pero confirmó la estructura descrita previamente.

#### 4.15. Purificación de XLF humana

Con el fin de estudiar este factor de reciente descubrimiento implicado en la reparación de DNA, el cual realiza sus funciones mediante la interacción con el complejo Xrcc4-DNA ligasa IV, se llevó a cabo la expresión y purificación de la proteína XLF humana recombinante. El vector de expresión utilizado fue pGEX6P1, el cual permitió la expresión del gen XLF unido a la proteína de fusión GST. En primer lugar se utilizaron las cepas de *E.coli* BL21(DE3), y se llevó a cabo un ensayo utilizando distintos tiempos de inducción, temperaturas, y concentraciones de IPTG. Tras la inducción de la expresión de la proteína de interés, se observó que la banda de la proteína sobreexpresada correspondía al peso molecular de XLF unida a la proteína de fusión GST (55 kDa en total) (Figura 4.25, A). En este ensayo se determinó que la expresión aumentaba con la temperatura, resultando ser ésta máxima a 37 °C. Sin embargo, cuando se llevó a cabo el ensayo de solubilidad, se observó que la mayor parte de la proteína expresada se encontraba insoluble (Figura 4.25, B y C). A continuación se llevó a cabo un segundo ensayo utilizando las condiciones de purificación previamente descritas (Yu et al., 2008) (Figura 4.25, D), pero al igual que en el caso anterior, aunque la proteína se sobreexpresó en grandes cantidades, la proteína no se expresaba de forma soluble (Figura 4.25, E).

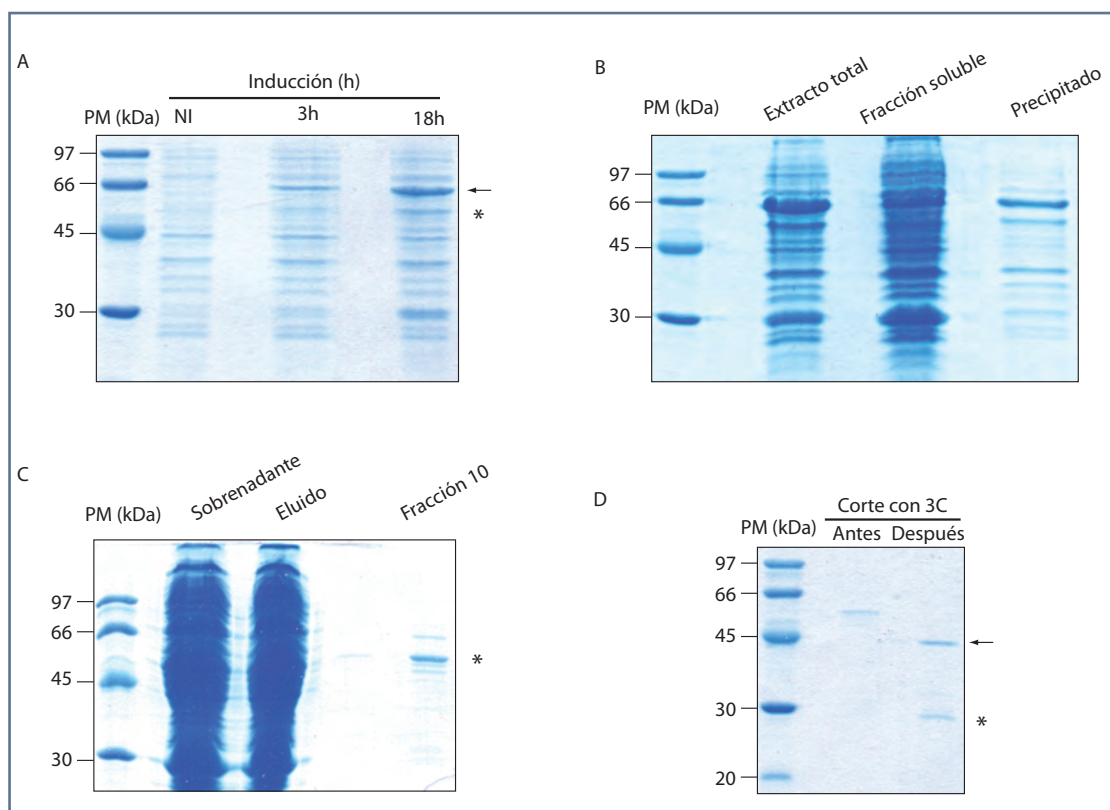
Se emplearon entonces nuevas estrategias de purificación. Se expresó y purificó XLF en cepas *E.coli* Rosetta2 mediante inducción con 0,25 mM IPTG a 15 °C durante toda la noche. Se observó que la expresión se había iniciado a las tres horas desde la inducción, pero ésta resultó máxima después de 18 horas. Sin embargo, sólo se produjo una



**Figura 4.25. Inducción y expresión de XLF-GST en células *E.coli* BL21.** **A)** Inducción de XLF-GST a distintas temperaturas (15, 22y 37 °C). **B-C)** Ensayo de solubilidad de la proteína expresada. **D)** Nueva inducción utilizando condiciones diferentes de tiempo y concentración de IPTG a las anteriores. **E)** Ensayo de solubilidad con las nuevas condiciones empleadas. (NI: No inducido; PM: marcadores de peso molecular).

inducción pequeña de la expresión de la proteína de interés (Figura 4.26, A, marcada con “\*”), mientras que otras proteínas de *E.coli* mostraron una mayor sobreexpresión (Figura 4.26, A, indicada con una flecha). En el ensayo de solubilidad realizado posteriormente se determinó que sólo una pequeña fracción de la proteína de interés expresada permanecía soluble (Figura 4.26, B).

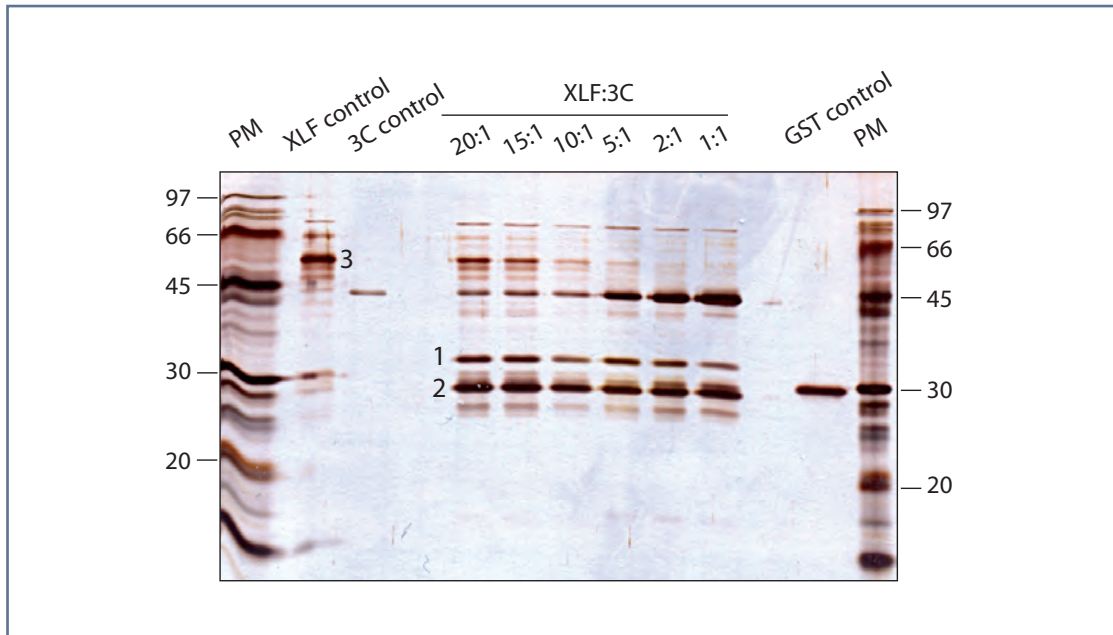
A pesar del bajo rendimiento obtenido, se llevó a cabo la purificación de la proteína contenida en la fracción soluble en una resina de glutatión sefarosa, donde se purificó una proteína con un peso molecular aparente correspondiente al complejo constituido por XLF y la proteína de fusión GST (Figura 4.26, C, marcada con “\*”). Se eluyó con glutatión y tras un proceso de diálisis se incubó con la proteasa 3C durante toda la noche con el objetivo de separar la proteína GST de XLF. Las especies resultantes del corte con la proteasa fueron analizadas mediante electroforesis en condiciones desnaturalizantes (Figura 4.26, D), en la que se observaron dos bandas que deberían corresponder, según los marcadores de peso molecular a la proteasa 3C unida a GST (~45 kDa, marcada con una flecha en la figura) y GST (~28 kDa, marcada con “\*” en la figura). Sin embargo, no se observó ninguna banda que pudiera corresponder según el



**Figura 4.26. Inducción y expresión de XLF-GST en células *E. coli* Rosetta. A)** Se tomaron muestras tras la inducción de XLF-GST a distintos tiempos (0, 3 y 18 h). **B)** Ensayo de solubilidad de la proteína expresada. **C)** Gel SDS-PAGE de las fracciones resultantes de la purificación mediante una resina de glutatión sefarosa. **D)** Gel SDS-PAGE en el que se analizaron los productos obtenidos tras el corte con la proteasa 3C para separar la proteína XLF de la GST.

peso molecular de los marcadores a la proteína XLF (~33 kDa).

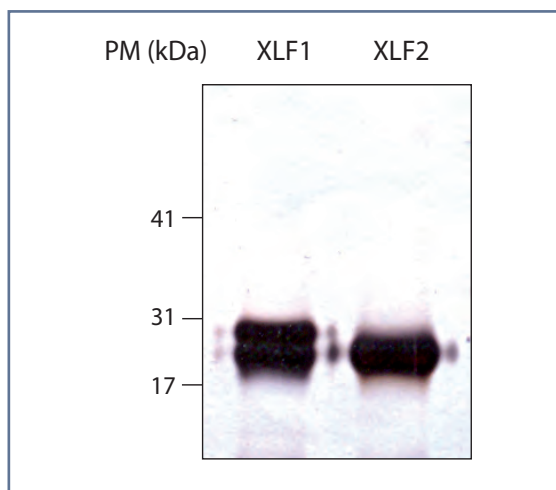
Se llevó a cabo a continuación una segunda prueba de corte con la proteasa 3C, empleando en este caso distintas relaciones proteína/proteasa (Figura 4.27) y se obtuvieron diversas bandas tras el corte, siendo dos de ellas mayoritarias. Estas bandas presentaban ligeras diferencias en la movilidad electroforética, de alrededor de 30 kDa (marcadas con “\*”). Se llevó a cabo un análisis mediante espectrometría de masas, donde se determinó que la banda de mayor movilidad electroforética contenía simultáneamente GST y XLF. Además, la banda de menor movilidad electroforética correspondía también a XLF. Estos análisis parecían indicar que la proteína XLF sufría una degradación durante el proceso de purificación hacia fragmentos de menor peso molecular. Por este motivo, se generaron dos construcciones de XLF que eran estables a lo largo del proceso de purificación, y que se encontraban delecionadas en su extremo C-terminal, denominadas XLF1 (XLFΔ248-299) y XLF2 (XLFΔ242-299), las cuales fueron purificadas siguiendo el mismo protocolo que para XLF silvestre. En el caso de XLF1, ésta continuaba sufriendo degradación parcial hasta generar la versión truncada XLF2, que parece ser la más estable de las dos formas (Figura 4.28).



**Figura 4.27. Ensayo de corte con la proteasa 3C y análisis mediante espectrometría de masas.** Se utilizaron distintas relaciones proteína/proteasa y se incubaron durante toda la noche. Los productos resultantes del corte con la proteasa 3C se analizaron mediante espectrometría de masas, así como el producto inicial previo al corte. La banda 1 se determinó que correspondía a XLF, la banda 2 contenía simultáneamente XLF y GST tras el corte y la banda 3 correspondía a la proteína XLF unida a la proteína de fusión GST antes del corte.

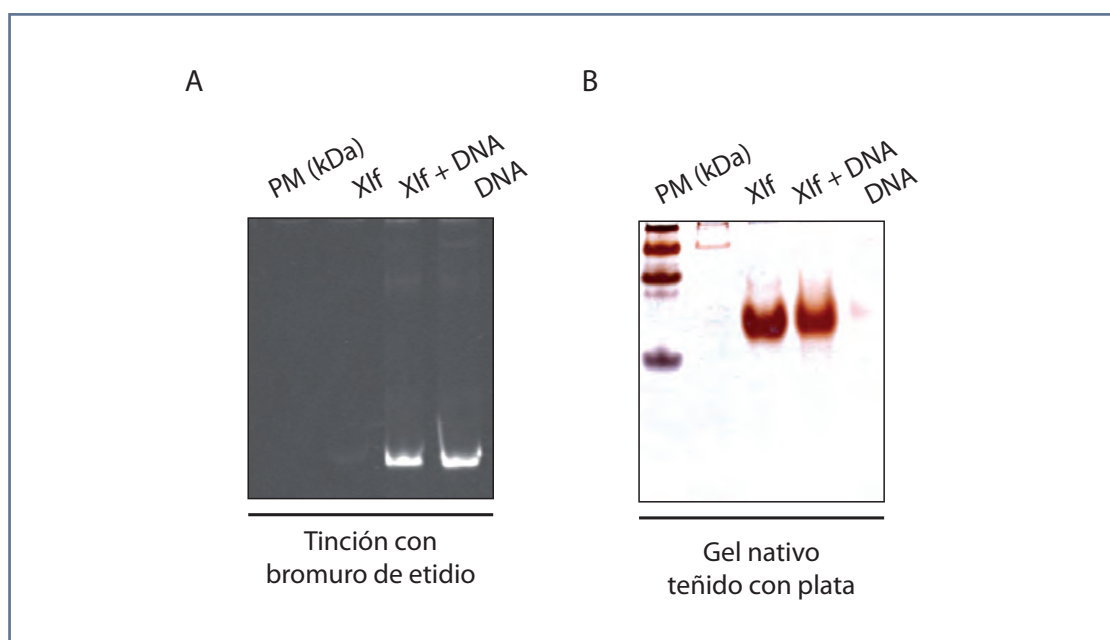
#### 4.16. Análisis de la interacción XLF-DNA mediante geles nativos

Con el fin de analizar la posible interacción entre la proteína XLF y DNA se realizaron ensayos de retardo en gel en condiciones nativas. Los ensayos de unión a DNA se llevaron a cabo con el fragmento de DNA de 390 pb descrito en Materiales y Métodos y con una de las formas truncadas de XLF purificadas (XLF2) en una relación molar 20:1. Se resolvieron los complejos en geles de poliacrilamida en condiciones nativas (Figura 4.29), los cuales fueron teñidos en primer lugar con bromuro de etidio para detectar el DNA y posteriormente con plata para detectar XLF. En este ensayo se determinó que la



**Figura 4.28. Gel SDS-PAGE en el que se analizaron las proteínas XLF truncadas purificadas.** XLF1 corresponde a la construcción XLF $\Delta$ 248-299, mientras que XLF2 corresponde a la construcción XLF $\Delta$ 242-299. En el caso de XLF1, se puede observar como esta proteína continuaba sufriendo degradación parcial una vez purificada, hasta generar la versión truncada estable XLF $\Delta$ 242-299 o XLF2.





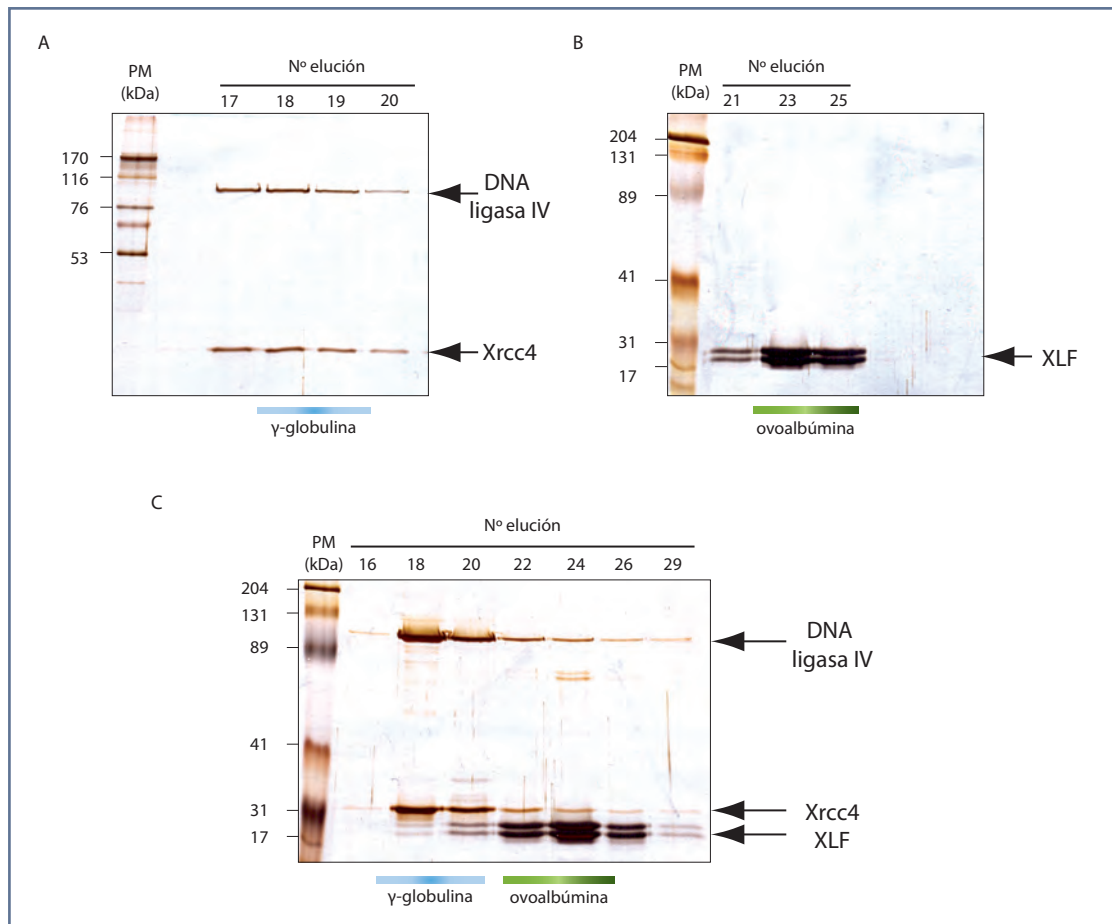
**Figura 4.29. Ensayo de interacción entre XLF y DNA.** **A)** Gel nativo teñido con bromuro de etidio, en el que se observa que no hay retardo en el DNA. **B)** El mismo gel nativo fue teñido posteriormente con plata, donde se observó que la movilidad electroforética tanto en presencia como en ausencia del DNA era la misma.

forma de XLF que se encuentra truncada en su extremo C-terminal no une DNA.

#### 4.17. Caracterización de la interacción entre XLF y el complejo Xrcc4-DNA ligasa IV

Se realizaron diversos ensayos de unión entre las proteínas XLF1 y el complejo Xrcc4( $\Delta$ 214-336)/DNA ligasa IV (se utilizaron estas especies porque disponíamos de mayor cantidad de proteína en ambos casos). Se incubó el complejo Xrcc4-DNA ligasa IV con la proteína XLF en una relación molar 1:3. A continuación se llevó a cabo una cromatografía de exclusión molecular en una columna Superdex 200 (Figura 4.30, C), inyectando como control las mismas cantidades de proteína de manera independiente que las utilizadas en el ensayo de unión (Figura 4.30, A y B). El cromatograma obtenido mostró que la proteína eluía en dos picos, que cuando fueron analizados mediante electroforesis en condiciones desnaturalizantes, mostraron que no se producía interacción entre ambas proteínas, puesto que XLF no co-eluyó en la misma fracción que uno de los componentes de mayor peso molecular correspondiente al complejo Xrcc4-DNA ligasa IV.

También se llevaron a cabo ensayos de interacción mediante geles nativos, tanto con el complejo Xrcc4-DNA ligasa IV truncado como silvestre, utilizando distintas condiciones, pero en todos los casos se obtuvieron resultados similares. Estos resultados negativos tuvieron sentido posteriormente a la luz de estudios publicados por otros



**Figura 4.30. Ensayo de unión entre el complejo Xrcc4( $\Delta$ 214-336)-DNA ligasa IV y XLF1. A)** Control del complejo Xrcc4-DNA ligasa IV. El complejo eluye en la misma fracción que el marcador  $\gamma$  globulina (158 kDa). **B)** Control de XLF1 (XLF $\Delta$ 248-299), el cual eluye entre los marcadores  $\gamma$  globulina (158 kDa) y ovoalbúmina (44 kDa). **C)** Complejo Xrcc4-DNA ligasa IV incubado con XLF1.

grupos en los que se describió que estas proteínas interaccionan débilmente (Andres et al., 2007; Deshpande y Wilson, 2007; Li et al., 2008). *In vivo*, se ha observado que el complejo Xrcc4-DNA ligasa IV y XLF son reclutados independientemente al sitio del daño, mediado por el heterodímero Ku (Yano et al., 2008), por lo que es posible que la formación de un complejo ternario sólo sea posible en el contexto de un complejo que contenga tanto las proteínas Ku como DNA.







## DISCUSIÓN



Los daños en el DNA son reparados por una multitud de complejos macromoleculares de gran tamaño constituidos por un amplio número de proteínas que actúan de forma dinámica sobre el DNA dañado, ensamblándose y desensamblándose a lo largo del ciclo celular (Weterings y van Gent, 2004). Esta participación dinámica de moléculas también implica que éstas adquieran distintas conformaciones y funciones a lo largo del ciclo celular para regular su actividad, la de otras moléculas con las que interaccionan, o esas mismas interacciones.

Para poder comprender los fundamentos moleculares de las lesiones en el DNA y su mecanismo de reparación es necesario resolver la estructura tridimensional de estos grandes complejos macromoleculares. La determinación estructural mediante microscopía electrónica es una herramienta muy útil en estos casos, puesto que permite el estudio de muestras biológicas con una pequeña cantidad del espécimen en cuestión. Otra ventaja es que permite estudiar muestras que presentan cambios conformacionales o son flexibles. Aunque la mayoría de las estructuras se resuelven a media-baja resolución, los avances tanto técnicos como computacionales están permitiendo alcanzar, con cada vez mayor frecuencia, altas resoluciones, incluso quasiatómicas en algunos casos (Jiang et al., 2008; Yu et al., 2008; Zhang et al., 2008).

En este contexto, el proceso de reparación conocido como NHEJ es llevado a cabo por macromoléculas que se localizan formando grandes complejos sobre el DNA, por lo que se trata de un buen material para ser estudiado mediante microscopía electrónica. Nuestro grupo ha llevado a cabo previamente la determinación estructural, mediante microscopía electrónica, de otras proteínas y complejos que intervienen en este proceso, como el heterodímero Ku (Rivera-Calzada et al., 2007). También se ha estudiado la proteína DNA-PKcs (Boskovic et al., 2003; Rivera-Calzada et al., 2005), y el complejo Ku/DNA-PKcs (Spagnolo et al., 2006). Siguiendo con esta línea de investigación, el siguiente objetivo propuesto en el grupo fue la determinación de la estructura del complejo Xrcc4-DNA ligasa IV, complejo que lleva a cabo la ligación del DNA durante el proceso de reparación. Las proteínas Xrcc4 y DNA ligasa IV forman un complejo *in vivo* que liga los extremos de DNA de doble cadena, producidos por diversas causas tanto externas como fisiológicas. La información estructural disponible en la actualidad para estas proteínas se limita a dominios aislados resueltos mediante otras técnicas estructurales.

Poco tiempo después del comienzo de estos estudios, se descubrió una nueva proteína que participaba en este proceso de reparación denominada Cernunnos/XLF. Se determinó que interaccionaba con el complejo Xrcc4-DNA ligasa IV, y por este motivo, se decidió llevar a cabo su expresión y purificación con el fin de formar un complejo estable con el complejo Xrcc4-DNA ligasa IV y realizar posteriormente su análisis estructural.

### 5.1. La estructura tridimensional de Xrcc4 completa revela una región carboxilo-terminal globular

Xrcc4 es una proteína esencial en el proceso de reparación mediante NHEJ en eucariotas y en la recombinación V(D)J en vertebrados. Xrcc4, como ya se ha comentado anteriormente, es fundamental para la estabilidad y la actividad de la proteína DNA ligasa IV (Grawunder et al., 1997; Modesti et al., 1999).

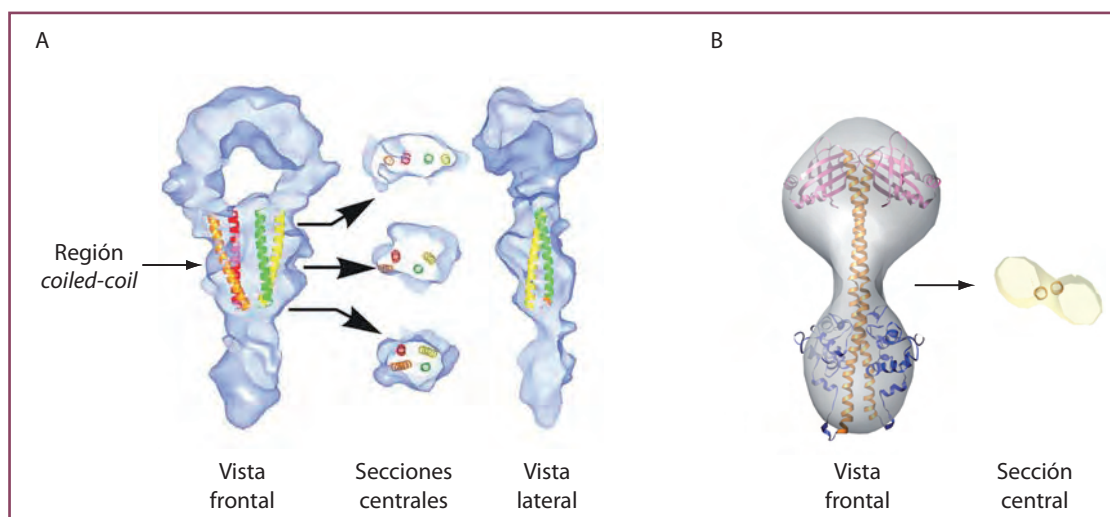
En esta tesis se ha llevado a cabo un estudio estructural mediante microscopía electrónica de la proteína Xrcc4. Este objetivo se planteó inicialmente porque no se conocía la estructura de la proteína completa, si bien existían estructuras parciales de la misma. Diversos estudios cristalográficos previos han proporcionado información estructural acerca de una versión truncada de Xrcc4 que carece aproximadamente de 1/3 de la molécula en su extremo C-terminal (Junop et al., 2000; Sibanda et al., 2001; Wu et al., 2009), dando lugar a un fragmento más estable para su cristalización. Por otro lado, la resolución de la estructura de la proteína completa podría resultar muy útil para llevar a cabo la interpretación de la estructura del complejo Xrcc4-DNA ligasa IV estudiado posteriormente.

Al intentar purificar la proteína Xrcc4 humana se observó que ésta era bioquímicamente inestable, determinándose que formaba oligómeros de alto peso molecular debido a la formación de puentes disulfuro intermoleculares. En el caso de la resolución de la estructura atómica de Xrcc4 unida a un péptido de la proteína DNA ligasa IV (Sibanda et al., 2001), los autores afirman que mutaron todos los residuos de cisteínas de Xrcc4 (cuatro en total, 93, 128, 130 y 165) a alaninas con el fin de prevenir la formación de oligómeros de Xrcc4 debidos a estos puentes disulfuro. Por este motivo, se decidió purificar la proteína de ratón, la cual no presentaba este comportamiento y, como se ha comentado anteriormente, presenta una elevada homología con la proteína de origen humano. En estudios realizados por otros grupos se ha determinado que Xrcc4 interacciona con moléculas de DNA de más de 100 pb, incluso en ausencia de la ligasa (Modesti et al., 1999). En esta tesis también se ha determinado que la proteína interacciona con moléculas de DNA de gran longitud, en concreto con una molécula de 390 pb.

La microscopía electrónica de partículas individuales nos ha permitido observar por primera vez la arquitectura tridimensional de la proteína Xrcc4 de mamíferos completa. El dímero, como ya se ha comentado previamente, tiene un peso molecular de sólo 76 kDa, por lo que se encuentra en el límite de lo que es posible estudiar mediante microscopía electrónica (Figura 5.1, B). Los especímenes apropiados para su estudio mediante esta técnica suelen ser muestras de más de 200 kDa para tinción negativa y más de 300 kDa para criomicroscopía. Sin embargo, han sido resueltas otras moléculas de pequeño tamaño, como son la prefoldina eucariota, de tan sólo 90 kDa (Martin-

Benito et al., 2002) o la geminina humana (Okorokov et al., 2004), también con un peso molecular aproximado de 90 kDa (Figura 5.1, A).

A pesar de su pequeño tamaño, Xrcc4 se visualizaba perfectamente en el microscopio utilizando tinción negativa. Sin embargo, el primer problema que surgió fue que, a pesar de que la muestra se encontraba completamente pura, cuando era observada al microscopio se detectaba una gran variabilidad en cuanto a formas y tamaños. En un principio, se pensó que la muestra podía sufrir un proceso de degradación durante los ciclos de congelación/descongelación a los que era sometida durante su almacenaje. Por otro lado, el agente de tinción puede producir deformaciones y aplastamientos en el espécimen bajo estudio que podrían ser responsables de la variabilidad observada. Es muy importante que la muestra observada al microscopio sea lo más homogénea posible, puesto que cuando se lleva a cabo su reconstrucción tridimensional, como ya se ha comentado en el apartado de introducción, ésta se sustenta en que las moléculas que se observan en el microscopio proceden de la misma estructura y en el mismo estado conformacional y de oligomerización. A pesar de que existen opciones para obtener un conjunto de datos homogéneo de una forma computacional mediante programas que analizan la heterogeneidad en las partículas y en los volúmenes, esta capacidad de análisis es limitada, por lo que resulta siempre recomendable obtener inicialmente una muestra bioquímicamente homogénea. Por este motivo, realizamos un proceso de purificación adicional de la muestra mediante centrifugación en gradientes de glicerol previo a su observación en el microscopio electrónico, con la finalidad de aislar y purificar el complejo dimérico de posibles agregados y partículas degradadas. Tras este proceso, se observó que la muestra visualizada al microscopio era más homogénea. Las partículas que se podían observar eran alargadas, pero aún así las imágenes



**Figura 5.1. Comparación de la estructura de Xrcc4 con la molécula de geminina humana, otro ejemplo de molécula de pequeño tamaño. A)** Estructura de la geminina resuelta mediante microscopía electrónica, representada en distintas orientaciones (adaptada de Okorokov et al., 2004). **B)** Estructura de Xrcc4, orientación frontal y superior.

seleccionadas mostraban un aspecto más grueso de lo esperado a nivel de la región central, puesto que en esa región era donde se esperaba localizar las alfa-hélices de la molécula. Las medias bidimensionales obtenidas inicialmente también reflejaban una densidad central que, a priori, parecía difícil asignar a dos alfa-hélices, por lo que pensamos que se podrían tratar de tetrámeros en lugar de dímeros. Esta idea era fundada, puesto que varios grupos habían determinado tanto bioquímicamente como estructuralmente que existía esta forma tetramérica de la molécula (Junop et al., 2000; Leber et al., 1998), aunque posteriormente se ha determinado que los tetrámeros no constituyen una forma funcional *in vivo* (Modesti et al., 2003).

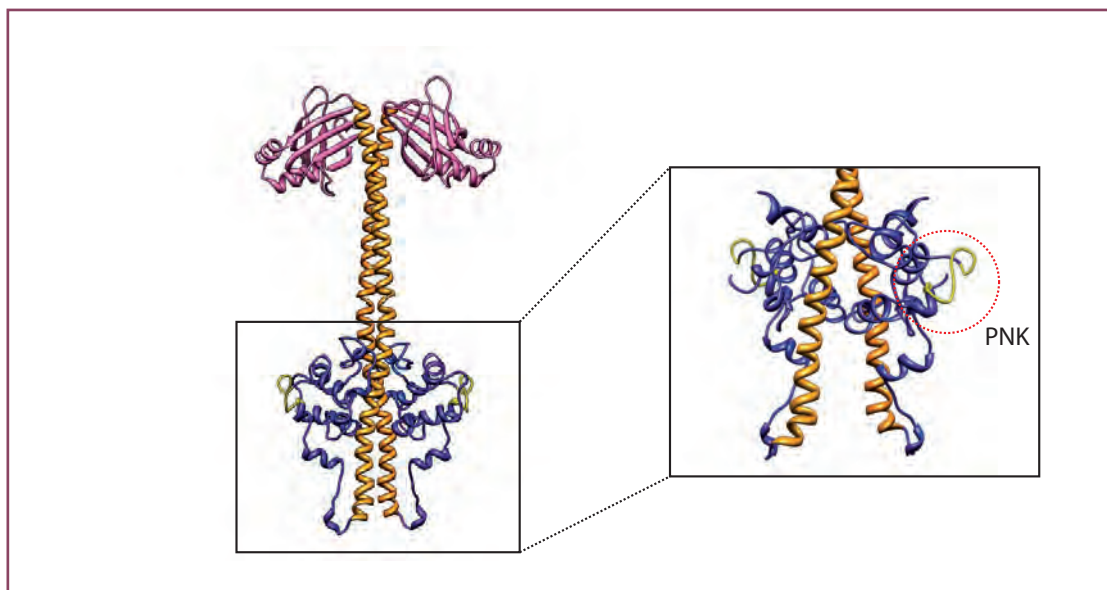
Una vez confirmamos mediante ensayos bioquímicos y biofísicos que la forma mayoritaria en nuestra preparación eran dímeros de Xrcc4, una posible explicación a este hecho era que las imágenes fueran más gruesas de lo esperado por el efecto de la tinción, de manera que esto podía dar cuenta de esa densidad adicional. Si comparamos nuestra estructura con la estructura obtenida mediante microscopía electrónica de la geminina humana mediante tinción negativa (Okorokov et al., 2004), una proteína de dimensiones parecidas a Xrcc4, podemos observar ciertas similitudes (Figura 5.1). Además de presentar un tamaño similar, esta proteína está compuesta por una región de alfa-hélices de alrededor de 170 Å de longitud. En la estructura resuelta mediante microscopía electrónica se puede observar que la región que los autores asignan a las alfa-hélices presenta una mayor densidad de la esperada para una región de alfa-hélices. Al igual que en nuestro caso, esto podría ser debido a un efecto artefactual de la tinción. Otra posible causa es que debido al pequeño tamaño de la partícula no dispongamos de la suficiente resolución para definir adecuadamente dos alfa-hélices mediante microscopía electrónica, de manera que lo que realmente observemos en la zona central sea el efecto debido a una relación señal-ruido muy baja, que visualmente se manifiesta en un engrosamiento de esa región. De hecho, en los volúmenes obtenidos, según el filtro que se aplicase al volumen, se podían ver conexiones por las regiones exteriores del volumen, por el interior o se perdía la conexión completamente. Esto sugiere que a la resolución de la estructura obtenida, 37 Å, se definen fundamentalmente los dos dominios globulares de los extremos, pero no es posible definir la masa central con claridad.

A pesar de la baja resolución obtenida, debida principalmente al pequeño tamaño de la molécula, la determinación de su estructura ha permitido proponer un modelo de organización, desconocido hasta la fecha, para la proteína completa. Dicha estructura presenta una longitud compatible con la molécula de Xrcc4, si se compara con la estructura dimérica cristalizada y resuelta de Xrcc4 humana (Junop et al., 2000). En el caso de la estructura resuelta mediante microscopía, Xrcc4 consta de dos regiones globulares en los extremos y una región central más estrecha. A pesar de la baja resolución, la forma característica de esta molécula ha permitido realizar un ajuste de la estructura atómica resuelta mediante cristalografía de rayos X de Xrcc4 humana.

Según este ajuste, la región N-terminal se localizaría en una de las zonas globulares y el dominio *coiled-coil* en la región central. Además, observábamos una región globular de gran tamaño, en la que no podíamos encajar ninguna estructura atómica y que debía corresponder al dominio C-terminal de Xrcc4. Como hemos comentado ya en otros apartados de esta tesis, se desconoce por completo la estructura tridimensional de esta región. Es importante porque contiene residuos que pueden ser fosforilados por DNA-PKcs y está filogenéticamente conservada (Yu et al., 2003), sugiriendo una posible función reguladora del mecanismo de NHEJ. De hecho, se ha postulado que la fosforilación de estas regiones por DNA-PKcs esté implicada en el desensamblaje del complejo del DNA (Merkle et al., 2002). Las predicciones de estructura secundaria sugerían que esta región podía estar compuesta principalmente por alfa-hélices y regiones desordenadas. Los experimentos de resonancia magnética nuclear llevados a cabo por nuestros colaboradores han determinado que este dominio presenta una estructura desordenada cuando se purifica de manera aislada, por lo que es probable que el resto de la estructura de Xrcc4 sea necesaria para su correcto plegamiento. En esta tesis hemos planteado que esta región C-terminal puede tratarse de un dominio globular localizado en el extremo opuesto al dominio N-terminal, algo que hemos visualizado por un lado en el microscopio electrónico y se ha apoyado por otro mediante la predicción de estructura terciaria realizada con Robetta. Hay que recordar que Xrcc4 no presenta homología con ninguna proteína conocida hasta el momento, y el modelado de la región C-terminal fue realizado mediante una aproximación *de novo*. Cuando llevamos a cabo el ajuste de la predicción con un mayor valor de correlación en el volumen de microscopía, observamos que ésta encajaba correctamente dentro de la densidad del volumen. (ver Figura 4.10 de Resultados). Un dato importante es que la predicción de estructura secundaria de la proteína XLF, cuya estructura es similar a la de Xrcc4, también propone que su dominio C-terminal (residuos 200-299) está formado por regiones de alfa-hélices y regiones desordenadas. Sin embargo, cuando se resolvió parte de su estructura (residuos 1-233) mediante cristalografía de rayos X (ver Figura 5.3, A) se observó que la región constituida por los residuos 200-233 formaba una estructura definida alrededor de las *coiled-coil* parecida a las estructuras predichas mediante Robetta para Xrcc4 (ver Figura 4.9 de Resultados y Figura 5.2).

Se ha determinado que a través de la región C-terminal de Xrcc4 también se producen interacciones con otras proteínas que intervienen en el ciclo de reparación, como la proteína Polinucleótido quinasa (Koch et al., 2004). En un estudio realizado por Bernstein y col. (Bernstein et al., 2005), en el que resolvieron la estructura cristalográfica de la proteína Polinucleótido quinasa (PNK) de mamíferos unida a un fosfopéptido de Xrcc4 de origen humano (Ac-YDES(pT)DEESEKK-CONH<sub>2</sub>) situado en el dominio C-terminal de la proteína, observaron que éste interaccionaba con el dominio FHA de la quinasa. PNK es una quinasa/fosfatasa que actúa en los extremos del DNA para que puedan ser ligados. En la estructura predicha por Robetta, esa región se encontraba en

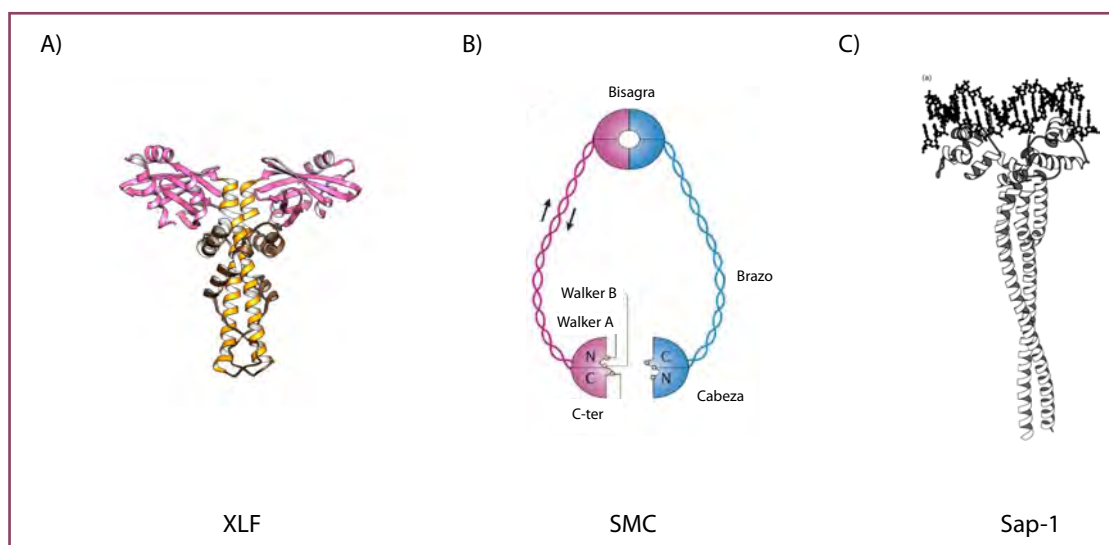




**Figura 5.2. Modelo de Robetta dimerizado obtenido para la proteína Xrcc4 completa.** Se muestra en amarillo la región de interacción de Xrcc4 con PNK.

una región que conecta dos alfa-hélices y se encontraría accesible para la interacción con la quinasa (Figura 5.2).

Existen otras proteínas con una estructura semejante a Xrcc4, basándose en sus regiones *coiled-coil*, y que están implicadas en interacciones con ácidos nucleicos o reparación de DNA. Uno de los últimos descubrimientos ha sido XLF (Andres et al., 2007; Li et al., 2008), el cual muestra ciertas similitudes estructurales con Xrcc4 a pesar de su escasa homología (Figura 5.3, A). Por otro lado, las proteínas SMC, encargadas del mantenimiento de la estructura cromosómica, y algunas de ellas implicadas en mecanismos de reparación de DNA, también presentan estructuras semejantes (Figura 5.3, B) (Hirano, 2006). Estos polipéptidos están constituidos por regiones con largas alfa-hélices estructuradas en dominios *coiled-coil* al igual que en Xrcc4, y también forman dímeros que interaccionan a través de regiones globulares. Estas estructuras presentan interacciones dinámicas con el DNA, manipulando las hebras y adoptando distintas conformaciones, más abiertas o cerradas gracias a la flexibilidad en la región *coiled-coil*. Otra similitud consiste en que las proteínas SMC podrían unirse al DNA a través de una región de cargas positivas situada en el dominio de dimerización. (Uhlmann y Hopfner, 2006). Sap 1 (del inglés *Switch-activating protein 1*), una proteína de la levadura *S. pombe*, con funciones en la organización y segregación cromosómica, también presenta una estructura similar. Es un dímero de 46 kDa que presenta una región *coiled-coil* de más de 120 Å de longitud (Bada et al., 2000). En esta proteína, se especula que el dominio N-terminal se puede unir a DNA, mientras que la región C-terminal puede servir para conectar con otras regiones o proteínas alejadas del DNA. Por tanto, parece ser que las regiones *coiled-coil* y su flexibilidad son necesarias para establecer las interacciones con otras proteínas y el DNA en estos grupos de proteínas.



**Figura 5.3. Representación esquemática y modelos de otras estructuras con dominios *coiled-coil*.** Existe un gran número de proteínas que presentan esta organización estructural y que están relacionadas con interacciones con ácidos nucleicos en mecanismos de reparación o mantenimiento de la integridad genómica. A) Estructura de XLF (Andres et al., 2007) (PDB 2R9A). B) Esquema de la organización de las proteínas SMC (adaptado de Hirano, 2006). C) Predicción de la estructura tridimensional de la proteína Sap1 en contacto con una molécula de DNA (tomado de Bada et al., 2000).

## 5.2. La región N-terminal de la proteína DNA-ligasa IV se localiza próxima a la región N-terminal de Xrcc4 en el complejo, mientras que la región C-terminal se localiza en el extremo opuesto

Una vez realizada la reconstrucción tridimensional de Xrcc4, procedimos a reconstruir el complejo de Xrcc4 con la proteína DNA ligasa IV, partiendo con la ventaja de que ya conocíamos la estructura de uno de sus componentes. Xrcc4-DNA ligasa IV es un complejo constitutivo y estable en la célula. La vida media de la ligasa se ha determinado entre 30 y 90 minutos en ausencia de Xrcc4, y es debido a esa inestabilidad por lo que no se puede purificar y estudiar independientemente. Cuando se iniciaron los estudios estructurales del complejo, solamente se conocía la estructura de dos terceras partes de Xrcc4 unida a un pequeño fragmento de uno de los dominios BRCT de la DNA ligasa IV, donde se determinó que una ligasa se une a un dímero de Xrcc4 de manera asimétrica (Sibanda et al., 2001). Por este motivo, resultaba interesante conocer la disposición y la interacción de estas proteínas dentro del complejo. A día de hoy, se han resuelto también mediante cristalografía de rayos X ambos dominios BRCT de la DNA ligasa IV, en complejo con el mismo fragmento de Xrcc4 resuelto de manera individual (Wu et al., 2009).

El análisis de las imágenes de microscopía electrónica del complejo ha proporcionado información estructural relevante sobre su organización. Las vistas laterales del complejo revelaron que éste estaba constituido por dos regiones globulares de mayor

densidad situadas en extremos opuestos, siendo un extremo más ancho y masivo y otro con forma de punta de flecha. Cuando se comparan estas imágenes con las de la estructura de Xrcc4 completa, se puede observar que la longitud de ambas moléculas es similar, pero en las medias del complejo se observa una región más ancha y otra más estrecha, a diferencia de las moléculas de Xrcc4, donde ambos extremos presentaban la misma anchura.

Para localizar ciertas regiones o dominios proteicos dentro de un complejo de mayor tamaño en las medias bidimensionales obtenidas mediante microscopía electrónica, existen distintas aproximaciones que son de uso rutinario. Entre las estrategias más comúnmente utilizadas destaca la generación de proteínas mutantes, de los que también se obtienen imágenes medias bidimensionales libres de referencia y su reconstrucción tridimensional, que son comparados posteriormente con los de la muestra silvestre. Tras realizar un mapa de diferencia, se puede determinar qué región se encuentra ausente en el mutante y, por tanto, asignarla a un determinado dominio proteico. Otras estrategias frecuentemente utilizadas consisten en emplear moléculas que se unan a determinadas regiones conocidas, como anticuerpos o moléculas de oro, de forma que se pueda identificar esa región (Frank, 2006). En general, la localización de las subunidades de un complejo de proteínas no es algo trivial, sobre todo en el caso de la tinción negativa, puesto que el agente de tinción limita la resolución alcanzada y dificulta en gran medida la asignación de los dominios sin ambigüedades. Sin embargo, en algunas ocasiones las medias bidimensionales libres de patrón obtenidas presentan suficiente grado de detalle de manera que no es necesario realizar la reconstrucción tridimensional para determinar la localización de un dominio (Nishida et al., 2006; Roberts et al., 2009). Durante esta tesis, se han llevado a cabo varias de estas estrategias con el fin de localizar determinados dominios de las proteínas dentro del complejo.

Por un lado, se ha localizado en el complejo la región N-terminal de la proteína DNA ligasa IV mediante un compuesto llamado Nanogold. Este compuesto contiene una pequeña molécula de oro unida a un átomo de níquel, el cual interacciona con las etiquetas de histidinas utilizadas de manera rutinaria en los procesos de purificación de proteínas. En el complejo Xrcc4-DNA ligasa IV, esta etiqueta estaba localizada en el extremo N-terminal de la ligasa. Tras purificar los complejos proteína-oro formados mediante una cromatografía de exclusión molecular, éstos se visualizaron en el microscopio electrónico. El procesamiento de las partículas con oro presentó ciertas dificultades, debido principalmente al valor de densidad extremo y constante de la partícula metálica. Éste dirigía e influenciaba el alineamiento de las imágenes, produciendo artefactos tanto en el alineamiento como en la reconstrucción tridimensional. A pesar de estos problemas, en algunas de las imágenes adquiridas se podía observar claramente una región de alta densidad a los electrones compatible en tamaño con la molécula de oro utilizada. Esta densidad se localizaba en las medias libres de patrón generadas en

la región más ancha del complejo Xrcc4-DNA ligasa IV.

Por otro lado, con el objetivo de localizar la posición del dominio C-terminal de la proteína Xrcc4 dentro del complejo humano, se generó un mutante que carecía de dicha región (Xrcc4( $\Delta$ 214-336)-DNA ligasa IV). A continuación, se visualizó al microscopio electrónico con el objetivo de su localización mediante un mapa de diferencia, como se ha comentado anteriormente. En las imágenes obtenidas, pero más claramente en las medias bidimensionales generadas, se observó que este complejo era mucho más pequeño en longitud que el silvestre, y se localizó el dominio C-terminal de Xrcc4 en la región que hemos denominado de punta de flecha dentro del complejo.

En resumen, mediante esta doble estrategia de marcaje se pudo definir que la zona N-terminal de Xrcc4 se encuentra próxima a la región catalítica de la ligasa (ver [Figura 5.4](#)).

### **5.3. La estructura tridimensional del complejo Xrcc4-DNA ligasa IV es consistente con los datos obtenidos mediante el análisis bidimensional previo**

Además de la localización de dominios en dos dimensiones, utilizamos las imágenes del complejo silvestre para llevar a cabo su reconstrucción tridimensional (ver [Figura 4.22 de Resultados](#)). Las estructuras obtenidas mediante los distintos procesamiento presentaban las mismas dimensiones, y todas ellas tenían una morfología similar: una región alargada con dos dominios globulares en los extremos que recordaba a la estructura de la proteína Xrcc4 obtenida mediante microscopía electrónica, en este caso uno mucho más ancho que el otro. En el centro de esta estructura siempre aparecía una región bilobulada, mientras que en la parte posterior del volumen aparecía de forma consistente una densidad generalmente plana. Estos datos concordaban con las medias bidimensionales obtenidas a partir de las imágenes individuales, puesto que se observaban unas partículas alargadas, con una base más estrecha, y la parte superior más ancha.

Cuando se llevó a cabo el encaje de las estructuras atómicas disponibles para el complejo Xrcc4-DNA ligasa IV (ver [Figura 4.23 de Resultados](#)), se observó que la estructura tridimensional obtenida, además de ser consistente con el mapeo realizado en dos dimensiones, también lo era con las estructuras atómicas. Sin embargo, la ausencia de una estabilización clara de la estructura durante el refinamiento impidió obtener una estructura definitiva del complejo Xrcc4-DNA ligasa IV. Por este motivo, llevamos a cabo distintas aproximaciones para intentar resolver los problemas del procesamiento. Por un lado, se intentó generar una estructura independiente mediante criomicroscopía. La tinción negativa limita la resolución del mapa de densidad obtenido mediante microscopía electrónica y puede llegar a introducir artefactos importantes en algunas ocasiones. Un estudio mediante criomicroscopía siempre es deseable, pero los resul-

tados en este caso dependen del tamaño de la proteína, principalmente. En nuestro caso, no fuimos capaces de observar la proteína debido al pequeño tamaño del complejo Xrcc4-DNA ligasa IV. Por este motivo, se intentó llevar a cabo una reconstrucción tridimensional del complejo mediante criotinción. En esta técnica, la muestra se preserva en su estado nativo al tiempo que un agente de tinción aumenta el contraste. Sin embargo, tampoco logramos en este caso observar la proteína mediante esta técnica.

Por otro lado, se intentó mejorar la calidad de los datos obtenidos inicialmente mediante tinción negativa llevando a cabo un ensayo bioquímico de reciente aplicación en microscopía electrónica conocido como GraFix y que consiste en someter a la muestra a un gradiente de densidad en glicerol a la vez que la muestra se va fijando a lo largo del gradiente con un agente de fijación, como el glutaraldehído. Con esta aproximación la calidad de los datos mejoró, y aunque existían aún regiones con alta variabilidad, la forma general de la molécula era consistente con los datos obtenidos previamente.

Todo ello, en conjunto, sugiere que a pesar de que la estructura tridimensional no convergió, los volúmenes pueden aproximarse a la estructura del complejo. Es posible que la falta de convergencia sea debida a cierta heterogeneidad estructural del complejo, causada por la flexibilidad estructural de la ligasa que puede ser necesaria para llevar a cabo la ligación de los extremos del DNA.

### **5.4. Modelo estructural del complejo Xrcc4-DNA ligasa IV e implicaciones fisiológicas**

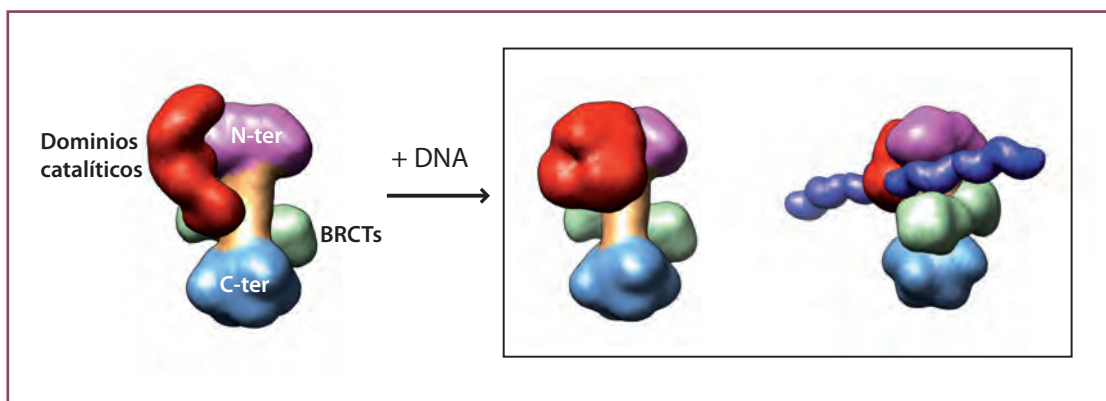
Los volúmenes obtenidos mediante reconstrucción de partículas individuales combinados con la localización tanto de la región N-terminal de la ligasa como de la C-terminal de Xrcc4, y las estructuras obtenidas previamente mediante cristalografía de rayos X para dominios aislados de Xrcc4 y la ligasa, nos permiten proponer un modelo de la arquitectura del complejo. La ligasa, como ya se ha mencionado anteriormente, comprende un núcleo catalítico en su extremo N-terminal seguido de los dominios BRCT. En nuestro modelo, los dominios BRCT rodearían la región *coiled-coil* de Xrcc4, mientras que el núcleo catalítico se localizaría en las proximidades del extremo N-terminal de Xrcc4 en su parte posterior. La posición de la región BRCT de la ligasa sugerida por la estructura tridimensional obtenida mediante microscopía electrónica en esta tesis implicaría que éstas podrían interaccionar con Xrcc4 a la vez que se encontrarían accesibles para la interacción con otras proteínas del ciclo de reparación. Entre las interacciones que ya se han descrito, la proteína Ku interacciona con la ligasa a través de su dominio BRCT1 (Costantini et al., 2007), y con Xrcc4, por lo que esa disposición de las proteínas del complejo podría ser coherente con las interacciones ya descritas. Además, todas estas interacciones se localizarían en el lado anterior del complejo mientras que el posterior quedaría libre para la interacción de la ligasa con el DNA. La posición y

orientación del dominio catalítico de la ligasa sugerido por la estructura implicaría que el DNA podría interaccionar con la región N-terminal de Xrcc4, a la vez que con el dominio catalítico de la ligasa, dada su proximidad. Aunque la interacción entre el dominio catalítico de la ligasa con Xrcc4 no ha sido observada previamente, sino que sólo se ha determinado la interacción de la región entre los dominios BRCT y el dominio BRCT2 con la región *coiled-coil* de Xrcc4, existe la posibilidad de que se requieran dominios adicionales para estabilizar la interacción entre las dos proteínas. De hecho, un estudio previo determinó que sólo era necesaria la región localizada entre ambos dominios BRCT para la interacción y estabilización del complejo, sin embargo, posteriormente se observó que el dominio BRCT2 también era necesario, por lo que no se puede descartar que existan otras interacciones entre ambas proteínas.

En cuanto a la unión del complejo a DNA, los estudios cristalográficos previos realizados con Xrcc4 han propuesto dos formas alternativas de unir DNA, aunque a día de hoy no existen datos bioquímicos o estructurales suficientes que confirmen ninguna de las dos teorías. Uno de los grupos sugiere que el DNA podría ser reconocido por el dominio globular situado en el extremo N-terminal de Xrcc4, debido al potencial electrostático positivo de esa región (Junop et al., 2000). En el segundo trabajo, se sugiere que el DNA podría interaccionar directamente con el dominio de alfa-hélices central (Sibanda et al., 2001). Si la interacción se produjera por la región N-terminal, Xrcc4 dispondría de una región de unión por cada monómero. Esta disposición facilitaría que cada subunidad de Xrcc4 interaccionase con uno de los dos fragmentos de DNA a reparar, estabilizando así los extremos para que puedan ser ligados por la ligasa, como sugieren Li y col. (Li et al., 1995) y Leber y col. (Leber et al., 1998).

Dado nuestro conocimiento actual de la reacción de ligación, se espera que la conformación de los dominios catalíticos de la ligasa cambien de manera significativa cuando se una al DNA durante el proceso de reparación (Figura 5.4). Las estructuras resueltas en otras ligasas, tanto eucariotas como bacterianas, han mostrado que las ligasas de distintos organismos rodean sus sustratos de una manera similar. Se ha determinado que el dominio *OB-fold* de las ligasas está unido al dominio de adenilación de una forma flexible y sufre una notable rotación entre el primer paso de adenilación de la enzima y los siguientes eventos del proceso de ligación (Shuman y Schwer, 1995). Esta forma de anillo puede aumentar la procesividad de la enzima y acelerar la búsqueda de la lesión a reparar (Ellenberger y Tomkinson, 2008), estabilizando así una conformación del DNA que exponga los extremos de la molécula que van a ser ligados al sitio activo de la enzima. Por tanto, las ligasas cambian de conformación durante la reacción de unión al DNA con el fin de acomodar las múltiples reacciones que ocurren con el nucleótido y los sustratos. La arquitectura de las ligasas con múltiples dominios proporciona la flexibilidad necesaria para ensamblarse y desensamblarse del DNA durante el proceso de reparación. La flexibilidad de este complejo queda patente por su capacidad de ligar extremos de DNA totalmente incompatibles (Gu et al., 2007).





**Figura 5.4. Modelo para la organización del complejo Xrcc4-DNA ligasa IV.** Se muestra en rojo los dominios catalíticos de la ligasa y en verde sus dominios BRCT. En rosa se muestra la región N-terminal de Xrcc4, en azul, su región C-terminal y en naranja su dominio coiled-coil. En presencia del DNA, es probable que ocurra un cambio conformacional a nivel de los dominios catalíticos de la ligasa que permita unirse al DNA con una mayor eficacia.

En este contexto, la estructura resuelta muestra que los dominios BRCT pueden proporcionar un punto de unión de la ligasa con Xrcc4 mientras que permite cierto grado de flexibilidad estructural de los dominios catalíticos. Esta libertad de movimiento sería necesaria para abrirse y cerrarse alrededor de la molécula de DNA, facilitando la ligación de las diversas configuraciones del DNA dañado. La capacidad para reparar el amplio rango de extremos dañados se refleja en la flexibilidad que todos los componentes de esta ruta tienen que presentar, tanto nucleasas, como polimerasas y ligasa. Xrcc4 también presenta flexibilidad, ya que se puede observar que las alfa-hélices resueltas en las diversas estructuras (ver Figura 1.8 de Introducción) adoptan una estructura curvada que es distinta en cada una de esas estructuras, lo que da una idea de su flexibilidad. Puede ser que esta cierta flexibilidad conformacional de los dominios catalíticos sea la responsable de las dificultades encontradas durante el procesamiento de las imágenes del complejo para determinar su estructura tridimensional.

### 5.5. Xrcc4 y DNA ligasa IV sólo interactúan de forma estable con DNA en presencia del heterodímero Ku

Se ha observado en trabajos anteriores que tanto Xrcc4 como el complejo Xrcc4-DNA ligasa IV interactúan con DNA (Chen et al., 2000; Modesti et al., 1999). En el caso de Xrcc4, la interacción se ha detectado únicamente con moléculas de DNA de más de 100 pb, de una manera no específica de secuencia (Modesti et al., 1999). Se ha observado que otras proteínas que participan en la reparación NHEJ también requieren moléculas de DNA relativamente largas para interactuar. Por ejemplo, XLF necesita una molécula de DNA de al menos 85 pb para unirse (Hentges et al., 2006; Lu et al., 2007). También se ha observado que en presencia de Ku, XLF y Xrcc4 se pueden unir a DNAs de una longitud inferior (Chen et al., 2000; Nick McElhinny et al., 2000;

Tsai et al., 2007).

Por otra parte, se ha determinado que la interacción entre Xrcc4 y el DNA se produce cuando la relación molar es mayor a 100:1, y que el complejo Xrcc4-DNA ligasa IV se une a DNA en unas condiciones en las que la relación molar es 20:1 (Nick McElhinny et al., 2000). Sin embargo, cuando Ku está presente en la reacción, estos valores disminuyen drásticamente, indicando que el complejo interacciona de una forma estable con el DNA exclusivamente en el contexto de un complejo ternario con Ku. En los ensayos de retardo realizados en esta tesis hemos comprobado que las proteínas unen DNA en las condiciones descritas previamente, y nunca cuando las concentraciones de proteína y DNA eran equimolares, por lo que en estas condiciones no es posible reconstituir un complejo con DNA para observarlo al microscopio electrónico. Será necesario por tanto ensamblar el complejo en presencia de la proteína Ku para reconstruirlo tridimensionalmente en presencia de DNA. El hecho de que Ku sea necesaria para formar un complejo estable de otras proteínas sobre el DNA tiene sentido, puesto que Ku es la proteína central en el proceso de reclutamiento y ensamblaje de todos los componentes del mecanismo de reparación NHEJ, incluido el complejo Xrcc4-DNA ligasa IV y XLF, sobre los extremos del DNA dañado (Nick McElhinny et al., 2000).

### **5.6. La región C-terminal de la proteína XLF introduce una inestabilidad que dificulta la purificación de la proteína completa**

Al poco tiempo de empezar a trabajar con el complejo Xrcc4-DNA ligasa IV se descubrió un nuevo factor denominado XLF ó Cernunnos, que interacciona con este complejo (Ahnesorg et al., 2006; Buck et al., 2006). Los objetivos del proyecto eran purificarlo y formar un complejo ternario con el complejo Xrcc4-DNA ligasa IV para determinar su estructura. Se ha determinado que XLF estimula a la ligasa durante la ligación de extremos desapareados y no cohesivos, de manera que puede regular la actividad del complejo en determinadas situaciones. *In vivo*, este hecho tiene relevancia, puesto que durante la recombinación V(D)J se producen extremos de DNA con extremos protuberantes desapareados en 3', y es en estos casos cuando aumenta la eficacia en el proceso de ligación gracias a Cernunnos. Parece que cuando XLF está presente, se mantiene ese extremo protuberante, manteniendo así la integridad de la secuencia del DNA. Los autores proponen que esa hebra continua puede servir de molde para que la polimerasa  $\mu$  polimerice la hebra complementaria (Gu et al., 2007; Li et al., 2008).

Cuando iniciamos su estudio, no se conocía la estructura de este factor. Tras los intentos de purificación infructuosos, conseguimos purificar la proteína XLF trunca en su extremo C-terminal, construcción de la que posteriormente fue resuelta su estructura tridimensional por dos grupos diferentes, (Li et al., 2008) y (Andres et al., 2007). Curiosamente, la estructura de XLF era muy parecida a la de Xrcc4, a pesar de



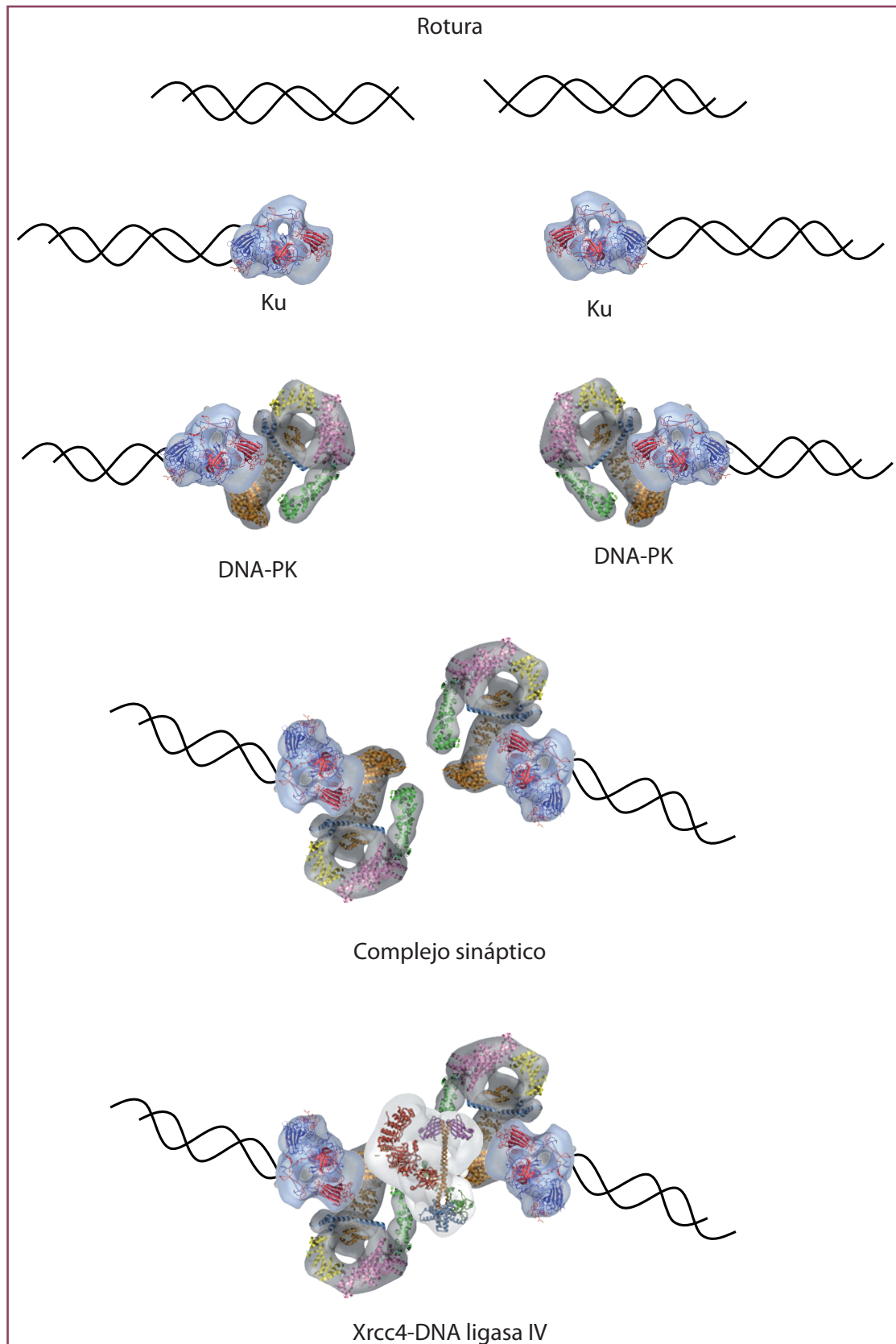


su baja homología de secuencia. Llevamos a cabo ensayos de unión a DNA, pero no se observó retardo en los ensayos de movilidad electroforética, lo cual pudimos explicar posteriormente porque un estudio reveló que XLF interacciona con el DNA a través de los últimos 75 residuos de su extremo C-terminal (Yano et al., 2008). En la estructura purificada por nosotros estaba troncada parte de esta región y, por este motivo, no observamos interacción con el DNA. Además, también intentamos realizar ensayos de unión de XLF con el complejo Xrcc4-DNA ligasa IV, pero en las condiciones utilizadas tampoco se observó interacción. *In vivo*, se ha visto que XLF es reclutado al sitio de la lesión de una forma rápida tras producirse el daño en el DNA, a través de Ku, pero de manera independiente al complejo Xrcc4-DNA ligasa IV (Yano et al., 2008). *In vitro*, se ha determinado que XLF interacciona con Xrcc4 a través de sus dominios N-terminal }, y estimula la actividad de la ligasa (Andres et al., 2007; Deshpande y Wilson, 2007; Li et al., 2008). Se ha determinado que la ligasa puede interaccionar con XLF a través de sus dominios BRCT, aunque débilmente. Es posible que estas proteínas interaccionen débilmente en ausencia de DNA y Ku y, por este motivo, no hayamos observado interacción. Será necesario llevar a cabo nuevos estudios para continuar caracterizando estructural y funcionalmente esta proteína.

## 5.7. El proceso de ligación en el contexto del NHEJ

A lo largo de los últimos años, en nuestro laboratorio se han resuelto las estructuras de diversos complejos que intervienen en la reparación de dobles roturas en la doble cadena de DNA mediante NHEJ. Toda esta información disponible, junto con los resultados obtenidos en esta tesis, se pueden integrar en un modelo para este mecanismo de reparación (Figura 5.5).

Los estudios estructurales del heterodímero Ku llevados a cabo mediante microscopía electrónica han mostrado que esta proteína experimenta cambios conformacionales en las regiones C-terminal de ambos monómeros al interaccionar con el DNA (Rivera-Calzada et al., 2007). Ku recluta a la proteína quinasa DNA-PKcs y a continuación, se desplaza hacia el interior de la molécula de DNA, dejando expuestos los dos extremos del DNA que favorecen el reclutamiento de DNA-PKcs. DNA-PKcs sufre cambios conformacionales cuando se une a Ku y al DNA, cuyo resultado es la activación de la actividad quinasa y la fosforilación por DNA-PKcs de otras proteínas del complejo de reparación. El resultado final es la formación de una holoenzima funcional denominada DNA-PK en el extremo de la rotura, la cual se visualizó por primera vez mediante microscopía electrónica (Spagnolo et al., 2006). Dos moléculas de DNA-PK situadas en dos moléculas distintas de DNA son capaces de formar complejos sinápticos en los extremos de la rotura, de forma que se produce así la aproximación de dos moléculas de DNA completamente desconectadas. Estos complejos se han visualizado mediante microscopía electrónica y se han resuelto también mediante SAXs (Hammel et al., 2010;



**Figura 5.5. Modelo del proceso de reparación mediante unión de extremos no homólogos.** En él se representan las estructuras que han sido resueltas mediante microscopía electrónica y el encaje en las mismas de las diferentes estructuras atómicas resueltas.

Rivera-Calzada et al., 2007). Esta disposición, así como la fosforilación por parte de DNA-PKcs de diversos sustratos, favorece el reclutamiento y ensamblaje de otras proteínas encargadas de procesar el DNA. Tras el procesamiento de los extremos, la ligación es llevada a cabo por el complejo Xrcc4-DNA ligasa IV, cuya actividad enzimática es estimulada gracias a interacciones proteína-proteína con la holoenzima DNA-PK (Chen et al., 2000; Nick McElhinny et al., 2000). En este contexto, la estructura del complejo Xrcc4-DNA ligasa IV determinada en esta tesis, sugiere que puede existir una interacción directa entre el núcleo catalítico de la ligasa y el dominio N-terminal del dímero de Xrcc4, mientras la región BRCT de la ligasa rodea el dominio *coiled-coil* de Xrcc4. En esta disposición, cuando el complejo es reclutado al sitio del daño, Xrcc4 podría unir el DNA mediante su dominio N-terminal a la vez que la ligasa interaccionaría con él a través de los dominios catalíticos. Xrcc4 actuaría como un punto de anclaje en el lugar de la lesión, interaccionando a la vez con DNA-PKcs, Ku, DNA ligasa IV y el DNA. Debido a su naturaleza dimerica, cada monómero de Xrcc4 se podría unir a un extremo del DNA, a la vez que interacciona con uno de los dos complejos sinápticos DNA-PK localizados en cada extremo del DNA. Al unirse al DNA y al resto de las proteínas del mecanismo de reparación, tienen que producirse cambios conformacionales en el complejo de manera que la región catalítica de la ligasa se cierre en torno al DNA para repararlo.

La reparación de las roturas producidas en la doble cadena de DNA no es un proceso sencillo, puesto que una vez ligada una de las hebras, es necesario reparar la segunda hebra para concluir el proceso. Hasta la fecha se desconoce si es un único complejo Xrcc4-DNA ligasa IV el que lleva a cabo la ligación de ambas hebras, con un proceso de adenilación intermedio, o se trata de dos complejos diferentes. Los resultados obtenidos mediante cristalografía de rayos X indican que la DNA ligasa se une de manera asimétrica a un dímero de Xrcc4, por lo que es posible que el proceso de ligación sea llevado a cabo por una única molécula de DNA ligasa IV por cada ciclo de NHEJ. Se ha demostrado que la ligasa puede actuar en una hebra del DNA de manera independiente de la otra (Gu et al., 2007; Ma et al., 2004; Ma et al., 2002), por lo que podría tratarse de una reacción secuencial de la misma ligasa. Aun así, no puede descartarse que otro complejo constituido por otro dímero de Xrcc4 y una ligasa lleven a cabo la ligación en la segunda hebra del DNA. En el caso de que intervenga una única ligasa, ésta proteína tiene que ser adenilada de nuevo tras la reparación de una de las hebras para proceder a la reparación de la segunda hebra. Desde un punto de vista biológico, esto implica que la proteína DNA ligasa IV puede ser recargada *in situ* para promover dos eventos de ligación. Sin embargo, este modelo tiene que ser apoyado por datos bioquímicos y no se puede descartar que puedan existir modelos alternativos en los que dos ligasas participen en un proceso de ligación.

Para seguir profundizando en el conocimiento del mecanismo de reparación mediante unión de extremos no homólogos es necesario resolver los interrogantes que aún existen en este proceso, como el papel de la autofosforilación de DNA-PKcs y la

fosforilación del resto de proteínas del complejo, o la caracterización de las interacciones entre el complejo Xrcc4-DNA ligasa IV con el DNA, entre otros. El objetivo final debe ser resolver la estructura y los mecanismos moleculares de un complejo de reparación completo, que incluya las proteínas DNA-PKcs, Ku70/80, XLF y Xrcc4-DNA ligasa IV sobre un DNA que mimetice una doble rotura en la cadena, así como los cambios conformacionales que medien la remodelación de la rotura, su reparación, y el ensamblaje final del complejo. En este sentido, la microscopía electrónica ha aportado información estructural de gran relevancia hasta el momento, y puede seguir siendo una herramienta fundamental, más ahora que se conocen un gran número de los componentes individuales que constituyen el complejo sobre el DNA. Sin embargo, existen otros procesos que no pueden ser estudiados mediante microscopía electrónica. Por tanto, la combinación entre estudios bioquímicos, funcionales y estructurales son la clave para el avance en el conocimiento de este mecanismo de reparación.



## CONCLUSIONES



- Se ha determinado mediante ensayos bioquímicos y biofísicos que Xrcc4 oligomeriza principalmente formando dímeros.
- Se ha propuesto un modelo tridimensional de la proteína Xrcc4 consistente en una molécula alargada donde los dominios amino y carboxi-terminal se encuentran en extremos opuestos conectados por la región *coiled-coil*.
- El dominio C-terminal de Xrcc4, que participa en la regulación de la reparación mediante unión de extremos no homólogos, consiste en un dominio globular constituido por alfa-hélices y un alto porcentaje de regiones desordenadas.
- La proteína DNA ligasa IV reconoce Xrcc4 mediante la interacción de sus dominios BRCT con la región *coiled-coil* de Xrcc4, mientras que la región catalítica de la ligasa se localiza en contacto con el dominio N-terminal de Xrcc4.
- El dominio C-terminal de Xrcc4 se localiza en el extremo del complejo opuesto al dominio catalítico de DNA ligasa IV
- La heterogeneidad observada en los datos del complejo Xrcc4-DNA ligasa IV puede ser un reflejo de una flexibilidad conformacional necesaria en el complejo para acomodarse a las necesidades estructurales de la ligación de una rotura.
- La purificación del factor XLF, recientemente implicado en la reparación NHEJ, muestra que la región C-terminal introduce una inestabilidad que dificulta la purificación de la proteína completa, si bien esta región es necesaria para el reconocimiento del DNA.
- No hemos observado una interacción estable entre el complejo Xrcc4-DNA ligasa IV y la proteína XLF. Es posible que esta interacción se vea favorecida en un contexto en el que estén presentes la proteína Ku y el DNA.





## **BIBLIOGRAFÍA**



- ♦ Abraham, R.T. (2004). PI 3-kinase related kinases: 'big' players in stress-induced signaling pathways. *DNA Repair (Amst)* 3, 883-887.
- ♦ Adrian, M., Dubochet, J., Fuller, S.D., y Harris, J.R. (1998). Cryo-negative staining. *Micron* 29, 145-160.
- ♦ Ahnesorg, P., Smith, P., y Jackson, S.P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* 124, 301-313.
- ♦ Andres, S.N., Modesti, M., Tsai, C.J., Chu, G., y Junop, M.S. (2007). Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. *Mol Cell* 28, 1093-1101.
- ♦ Aravind, L., y Koonin, E.V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* 11, 1365-1374.
- ♦ Bada, M., Walther, D., Arcangioli, B., Doniach, S., y Delarue, M. (2000). Solution structural studies and low-resolution model of the *Schizosaccharomyces pombe* sap1 protein. *J Mol Biol* 300, 563-574.
- ♦ Bernstein, N.K., Williams, R.S., Rakovszky, M.L., Cui, D., Green, R., Karimi-Busheri, F., Mani, R.S., Galicia, S., Koch, C.A., Cass, C.E., *et al.* (2005). The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Mol Cell* 17, 657-670.
- ♦ Bonneau, R., Strauss, C.E., Rohl, C.A., Chivian, D., Bradley, P., Malmstrom, L., Robertson, T., y Baker, D. (2002). De novo prediction of three-dimensional structures for major protein families. *J Mol Biol* 322, 65-78.
- ♦ Bork, P., Hofmann, K., Bucher, P., Neuwald, A.F., Altschul, S.F., y Koonin, E.V. (1997). A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J* 11, 68-76.
- ♦ Boskovic, J., Rivera-Calzada, A., Maman, J.D., Chacon, P., Willison, K.R., Pearl, L.H., y Llorca, O. (2003). Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. *Embo J* 22, 5875-5882.
- ♦ Branden, C., y Tooze, J. (1999). *Introduction to Protein Structure*, Second Edition edn (New York, Garland Publishing).
- ♦ Brnzei, D., y Foiani, M. (2008). Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol* 9, 297-308.
- ♦ Bryans, M., Valenzano, M.C., y Stamato, T.D. (1999). Absence of DNA ligase IV protein in XR-1 cells: evidence for stabilization by XRCC4. *Mutat Res* 433, 53-58.

- ♦ Buck, D., Malivert, L., de Chasseval, R., Barraud, A., Fondaneche, M.C., Sanal, O., Plebani, A., Stephan, J.L., Hufnagel, M., le Deist, F., *et al.* (2006). Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* 124, 287-299.
- ♦ Burma, S., Chen, B.P., y Chen, D.J. (2006). Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair (Amst)* 5, 1042-1048.
- ♦ Callebaut, I., Malivert, L., Fischer, A., Mornon, J.P., Revy, P., y de Villartay, J.P. (2006). Cernunnos interacts with the XRCC4/DNA-ligase IV complex and is homologous to the yeast nonhomologous end-joining factor NEJ1. *J Biol Chem*.
- ♦ Callebaut, I., y Mornon, J.P. (1997). From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. *FEBS Lett* 400, 25-30.
- ♦ Chan, D.W., Chen, B.P., Prithivirajasingh, S., Kurimasa, A., Story, M.D., Qin, J., y Chen, D.J. (2002). Autophosphorylation of the DNA-dependent protein kinase catalytic subunit is required for rejoining of DNA double-strand breaks. *Genes Dev* 16, 2333-2338.
- ♦ Chen, L., Trujillo, K., Sung, P., y Tomkinson, A.E. (2000). Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *J Biol Chem* 275, 26196-26205.
- ♦ Cheng, Y., y Walz, T. (2009). The advent of near-atomic resolution in single-particle electron microscopy. *Annu Rev Biochem* 78, 723-742.
- ♦ Chivian, D., y Baker, D. (2006). Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. *Nucleic Acids Res* 34, e112.
- ♦ Chivian, D., Kim, D.E., Malmstrom, L., Bradley, P., Robertson, T., Murphy, P., Strauss, C.E., Bonneau, R., Rohl, C.A., y Baker, D. (2003). Automated prediction of CASP-5 structures using the Robetta server. *Proteins* 53 Suppl 6, 524-533.
- ♦ Chivian, D., Kim, D.E., Malmstrom, L., Schonbrun, J., Rohl, C.A., y Baker, D. (2005). Prediction of CASP6 structures using automated Robetta protocols. *Proteins* 61 Suppl 7, 157-166.
- ♦ Cimprich, K.A., y Cortez, D. (2008). ATR: an essential regulator of genome integrity. *Nat Rev Mol Cell Biol* 9, 616-627.
- ♦ Costantini, S., Woodbine, L., Andreoli, L., Jeggo, P.A., y Vindigni, A. (2007). Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair (Amst)*.

- ♦ Costanzo, V., Shechter, D., Lupardus, P.J., Cimprich, K.A., Gottesman, M., y Gautier, J. (2003). An ATR- and Cdc7-dependent DNA damage checkpoint that inhibits initiation of DNA replication. *Mol Cell* 11, 203-213.
- ♦ Critchlow, S.E., Bowater, R.P., y Jackson, S.P. (1997). Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr Biol* 7, 588-598.
- ♦ Cromie, G.A., Connelly, J.C., y Leach, D.R. (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol Cell* 8, 1163-1174.
- ♦ DeFazio, L.G., Stansel, R.M., Griffith, J.D., y Chu, G. (2002). Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J* 21, 3192-3200.
- ♦ Deshpande, R.A., y Wilson, T.E. (2007). Modes of interaction among yeast Nej1, Lif1 and Dnl4 proteins and comparison to human XLF, XRCC4 and Lig4. *DNA Repair (Amst)* 6, 1507-1516.
- ♦ Doherty, A.J., Jackson, S.P., y Weller, G.R. (2001). Identification of bacterial homologues of the Ku DNA repair proteins. *FEBS Lett* 500, 186-188.
- ♦ Dore, A.S., Furnham, N., Davies, O.R., Sibanda, B.L., Chirgadze, D.Y., Jackson, S.P., Pellegrini, L., y Blundell, T.L. (2006). Structure of an Xrcc4-DNA ligase IV yeast ortholog complex reveals a novel BRCT interaction mode. *DNA Repair (Amst)* 5, 362-368.
- ♦ Downs, J.A., y Jackson, S.P. (2004). A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol* 5, 367-378.
- ♦ Drouet, J., Delteil, C., Lefrancois, J., Concannon, P., Salles, B., y Calsou, P. (2005). DNA-dependent protein kinase and XRCC4-DNA ligase IV mobilization in the cell in response to DNA double strand breaks. *J Biol Chem* 280, 7060-7069.
- ♦ Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowell, A.W., y Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* 21, 129-228.
- ♦ Ellenberger, T., y Tomkinson, A.E. (2008). Eukaryotic DNA ligases: structural and functional insights. *Annu Rev Biochem* 77, 313-338.
- ♦ Erickson, H.P. (2009). Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol Proced On-line*.
- ♦ Foster, R.E., Nnakwe, C., Woo, L., y Frank, K.M. (2006). Monoubiquitination of the nonhomologous end joining protein XRCC4. *Biochem Biophys Res Commun* 341, 175-

183.

- ♦ Frank, J. (2006). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, 2006 edn (New York, Oxford University Press).
- ♦ Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., y Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 116, 190-199.
- ♦ Garzon, J.I., Kovacs, J., Abagyan, R., y Chacon, P. (2007). ADP\_EM: fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinformatics* 23, 427-433.
- ♦ Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., y Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31, 3784-3788.
- ♦ Gómez-Moreno Calera, C., y Sancho Sanz, J. (2003). *Estructura de proteínas*, 2003 edn (Barcelona, Ariel).
- ♦ Gottlieb, T.M., y Jackson, S.P. (1993). The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* 72, 131-142.
- ♦ Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T.E., Mann, M., y Lieber, M.R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* 388, 492-495.
- ♦ Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., y Lieber, M.R. (1998a). DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Mol Cell* 2, 477-484.
- ♦ Grawunder, U., Zimmer, D., Kulesza, P., y Lieber, M.R. (1998b). Requirement for an interaction of XRCC4 with DNA ligase IV for wild-type V(D)J recombination and DNA double-strand break repair in vivo. *J Biol Chem* 273, 24708-24714.
- ♦ Grzesiek, S., y Sass, H.J. (2009). From biomolecular structure to functional understanding: new NMR developments narrow the gap. *Curr Opin Struct Biol* 19, 585-595.
- ♦ Gu, J., Lu, H., Tiffin, B., Shimazaki, N., Goodman, M.F., y Lieber, M.R. (2007a). XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J* 26, 1010-1023.
- ♦ Gu, J., Lu, H., Tsai, A.G., Schwarz, K., y Lieber, M.R. (2007b). Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Res* 35, 5755-5762.

- ♦ Hah, Y.S., Lee, J.H., y Kim, D.R. (2007). DNA-dependent protein kinase mediates V(D)J recombination via RAG2 phosphorylation. *J Biochem Mol Biol* 40, 432-438.
- ♦ Hainfeld, J.F., Liu, W., Halsey, C.M., Freimuth, P., y Powell, R.D. (1999). Ni-NTA-gold clusters target His-tagged proteins. *J Struct Biol* 127, 185-198.
- ♦ Hammel, M., Yu, Y., Mahaney, B.L., Cai, B., Ye, R., Phipps, B.M., Rambo, R.P., Hura, G.L., Pelikan, M., So, S., *et al.* (2010). Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. *J Biol Chem* 285, 1414-1423.
- ♦ Harrison, J.C., y Haber, J.E. (2006). Surviving the breakup: the DNA damage checkpoint. *Annu Rev Genet* 40, 209-235.
- ♦ Helleday, T., Lo, J., van Gent, D.C., y Engelward, B.P. (2007). DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA Repair (Amst)* 6, 923-935.
- ♦ Hentges, P., Ahnesorg, P., Pitcher, R.S., Bruce, C.K., Kysela, B., Green, A.J., Bianchi, J., Wilson, T.E., Jackson, S.P., y Doherty, A.J. (2006). Evolutionary and functional conservation of the DNA non-homologous end-joining protein, XLF/Cernunnos. *J Biol Chem* 281, 37517-37526.
- ♦ Heymann, J.B., y Belnap, D.M. (2007). Bsoft: image processing and molecular modeling for electron microscopy. *J Struct Biol* 157, 3-18.
- ♦ Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. *Nat Rev Mol Cell Biol* 7, 311-322.
- ♦ Ho, C.K., Van Etten, J.L., y Shuman, S. (1997). Characterization of an ATP-dependent DNA ligase encoded by Chlorella virus PBCV-1. *J Virol* 71, 1931-1937.
- ♦ Hoeijmakers, J.H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* 411, 366-374.
- ♦ Ilari, A., y Savino, C. (2008). Protein structure determination by x-ray crystallography. *Methods Mol Biol* 452, 63-87.
- ♦ Jackson, S.P. (2002). Sensing and repairing DNA double-strand breaks. *Carcinogenesis* 23, 687-696.
- ♦ Jiang, W., Baker, M.L., Jakana, J., Weigele, P.R., King, J., y Chiu, W. (2008). Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. *Nature* 451, 1130-1134.



- ♦ Ionic, S., Sorzano, C.O., y Boisset, N. (2008). Comparison of single-particle analysis and electron tomography approaches: an overview. *J Microsc* 232, 562-579.
- ♦ Junop, M.S., Modesti, M., Guarne, A., Ghirlando, R., Gellert, M., y Yang, W. (2000). Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. *Embo J* 19, 5962-5970.
- ♦ Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., *et al.* (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods* 5, 53-55.
- ♦ Kelley, L.A., y Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4, 363-371.
- ♦ Kelly, D.F., Abeyrathne, P.D., Dukovski, D., y Walz, T. (2008a). The Affinity Grid: a pre-fabricated EM grid for monolayer purification. *J Mol Biol* 382, 423-433.
- ♦ Kelly, D.F., Dukovski, D., y Walz, T. (2008b). Monolayer purification: a rapid method for isolating protein complexes for single-particle electron microscopy. *Proc Natl Acad Sci U S A* 105, 4703-4708.
- ♦ Kim, D.E., Chivian, D., y Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 32, W526-531.
- ♦ Koch, C.A., Agyei, R., Galicia, S., Metalnikov, P., O'Donnell, P., Starostine, A., Weinfeld, M., y Durocher, D. (2004). Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *Embo J* 23, 3874-3885.
- ♦ Kurimasa, A., Kumano, S., Boubnov, N.V., Story, M.D., Tung, C.S., Peterson, S.R., y Chen, D.J. (1999). Requirement for the kinase activity of human DNA-dependent protein kinase catalytic subunit in DNA strand break rejoining. *Mol Cell Biol* 19, 3877-3884.
- ♦ Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948.
- ♦ Leber, R., Wise, T.W., Mizuta, R., y Meek, K. (1998). The XRCC4 gene product is a target for and interacts with the DNA-dependent protein kinase. *J Biol Chem* 273, 1794-1801.
- ♦ Lee, J.H., y Paull, T.T. (2007). Activation and regulation of ATM kinase activity in response to DNA double-strand breaks. *Oncogene* 26, 7741-7748.
- ♦ Lee, J.W., Yannone, S.M., Chen, D.J., y Povirk, L.F. (2003). Requirement for XRCC4 and DNA ligase IV in alignment-based gap filling for nonhomologous DNA end joining

in vitro. *Cancer Res* 63, 22-24.

- ♦ Lee, K.J., Jovanovic, M., Udayakumar, D., Bladen, C.L., y Dynan, W.S. (2004). Identification of DNA-PKcs phosphorylation sites in XRCC4 and effects of mutations at these sites on DNA end joining in a cell-free system. *DNA Repair (Amst)* 3, 267-276.
- ♦ Leschziner, A.E., y Nogales, E. (2007). Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions. *Annu Rev Biophys Biomol Struct* 36, 43-62.
- ♦ Li, Y., Chirgadze, D.Y., Bolanos-Garcia, V.M., Sibanda, B.L., Davies, O.R., Ahnesorg, P., Jackson, S.P., y Blundell, T.L. (2008). Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. *EMBO J* 27, 290-300.
- ♦ Li, Z., Otevrel, T., Gao, Y., Cheng, H.L., Seed, B., Stamato, T.D., Taccioli, G.E., y Alt, F.W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* 83, 1079-1089.
- ♦ Lieber, M.R. (2008). The mechanism of human nonhomologous DNA end joining. *J Biol Chem* 283, 1-5.
- ♦ Lieber, M.R., Yu, K., y Raghavan, S.C. (2006). Roles of nonhomologous DNA end joining, V(D)J recombination, and class switch recombination in chromosomal translocations. *DNA Repair (Amst)* 5, 1234-1245.
- ♦ Llorca, O. (2005). Introduction to 3D reconstruction of macromolecules using single particle electron microscopy. *Acta Pharmacol Sin* 26, 1153-1164.
- ♦ Lobrich, M., y Jeggo, P.A. (2007). The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. *Nat Rev Cancer* 7, 861-869.
- ♦ Lovejoy, C.A., y Cortez, D. (2009). Common mechanisms of PIKK regulation. *DNA Repair (Amst)* 8, 1004-1008.
- ♦ Lu, H., Pannicke, U., Schwarz, K., y Lieber, M.R. (2007). Length-dependent binding of human XLF to DNA and stimulation of XRCC4.DNA ligase IV activity. *J Biol Chem* 282, 11155-11162.
- ♦ Lucic, V., Forster, F., y Baumeister, W. (2005). Structural studies by electron tomography: from cells to molecules. *Annu Rev Biochem* 74, 833-865.
- ♦ Ludtke, S.J., Baker, M.L., Chen, D.H., Song, J.L., Chuang, D.T., y Chiu, W. (2008). De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure* 16, 441-448.

- ♦ Ludtke, S.J., Baldwin, P.R., y Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128, 82-97.
- ♦ Ma, Y., Lu, H., Tippin, B., Goodman, M.F., Shimazaki, N., Koiwai, O., Hsieh, C.L., Schwarz, K., y Lieber, M.R. (2004). A biochemically defined system for mammalian non-homologous DNA end joining. *Mol Cell* 16, 701-713.
- ♦ Ma, Y., Pannicke, U., Schwarz, K., y Lieber, M.R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* 108, 781-794.
- ♦ Mahaney, B.L., Meek, K., y Lees-Miller, S.P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 417, 639-650.
- ♦ Manke, I.A., Lowery, D.M., Nguyen, A., y Yaffe, M.B. (2003). BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science* 302, 636-639.
- ♦ Marabini, R., Masegosa, I.M., San Martín, M.C., Marco, S., Fernández, J.J., de la Fraga, L.G., Vaquerizo, C., y Carazo, J.M. (1996). Xmipp: An Image Processing Package for Electron Microscopy. *J Struct Biol* 116, 237-240.
- ♦ Mari, P.O., Florea, B.I., Persengiev, S.P., Verkaik, N.S., Bruggenwirth, H.T., Modesti, M., Giglia-Mari, G., Bezstarosti, K., Demmers, J.A., Luidert, T.M., *et al.* (2006). Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc Natl Acad Sci U S A* 103, 18597-18602.
- ♦ Martín-Benito, J., Boskovic, J., Gómez-Puertas, P., Carrascosa, J.L., Simons, C.T., Lewis, S.A., Bartolini, F., Cowan, N.J., y Valpuesta, J.M. (2002). Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. *EMBO J* 21, 6377-6386.
- ♦ Matsumoto, Y., Suzuki, N., Namba, N., Umeda, N., Ma, X.J., Morita, A., Tomita, M., Enomoto, A., Serizawa, S., Hirano, K., *et al.* (2000). Cleavage and phosphorylation of XRCC4 protein induced by X-irradiation. *FEBS Lett* 478, 67-71.
- ♦ Merkle, D., Douglas, P., Moorhead, G.B., Leonenko, Z., Yu, Y., Cramb, D., Bazett-Jones, D.P., y Lees-Miller, S.P. (2002). The DNA-dependent protein kinase interacts with DNA to form a protein-DNA complex that is disrupted by phosphorylation. *Biochemistry* 41, 12706-12714.
- ♦ Mindell, J.A., y Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142, 334-347.
- ♦ Modesti, M., Hesse, J.E., y Gellert, M. (1999). DNA binding of Xrcc4 protein is as-

sociated with V(D)J recombination but not with stimulation of DNA ligase IV activity. *Embo J* 18, 2008-2018.

- ♦ Modesti, M., Junop, M.S., Ghirlando, R., van de Rakt, M., Gellert, M., Yang, W., y Kanaar, R. (2003). Tetramerization and DNA ligase IV interaction of the DNA double-strand break repair protein XRCC4 are mutually exclusive. *J Mol Biol* 334, 215-228.
- ♦ Mohammad, D.H., y Yaffe, M.B. (2009). 14-3-3 proteins, FHA domains and BRCT domains in the DNA damage response. *DNA Repair (Amst)* 8, 1009-1017.
- ♦ Neylon, C. (2008). Small angle neutron and X-ray scattering in structural biology: recent examples from the literature. *Eur Biophys J* 37, 531-541.
- ♦ Nick McElhinny, S.A., Snowden, C.M., McCarville, J., y Ramsden, D.A. (2000). Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol* 20, 2996-3003.
- ♦ Nickell, S., Kofler, C., Leis, A.P., y Baumeister, W. (2006). A visual approach to proteomics. *Nat Rev Mol Cell Biol* 7, 225-230.
- ♦ Nishida, N., Walz, T., y Springer, T.A. (2006). Structural transitions of complement component C3 and its activation products. *Proc Natl Acad Sci U S A* 103, 19737-19742.
- ♦ O'Driscoll, M., Cerosaletti, K.M., Girard, P.M., Dai, Y., Stumm, M., Kysela, B., Hirsch, B., Gennery, A., Palmer, S.E., Seidel, J., *et al.* (2001). DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Mol Cell* 8, 1175-1185.
- ♦ Odell, M., y Shuman, S. (1999). Footprinting of Chlorella virus DNA ligase bound at a nick in duplex DNA. *J Biol Chem* 274, 14032-14039.
- ♦ Odell, M., Sriskanda, V., Shuman, S., y Nikolov, D.B. (2000). Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining. *Mol Cell* 6, 1183-1193.
- ♦ Okorokov, A.L., Orlova, E.V., Kingsbury, S.R., Bagneris, C., Gohlke, U., Williams, G.H., y Stoeber, K. (2004). Molecular structure of human geminin. *Nat Struct Mol Biol* 11, 1021-1022.
- ♦ Pascal, J.M., O'Brien, P.J., Tomkinson, A.E., y Ellenberger, T. (2004). Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* 432, 473-478.
- ♦ Pascal, J.M., Tsodikov, O.V., Hura, G.L., Song, W., Cotner, E.A., Classen, S., Tomkinson, A.E., Tainer, J.A., y Ellenberger, T. (2006). A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Mol Cell* 24, 279-291.

- ♦ Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., y Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612.
- ♦ Povirk, L.F. (2006). Biochemical mechanisms of chromosomal translocations resulting from DNA double-strand breaks. *DNA Repair (Amst)* 5, 1199-1212.
- ♦ Riballo, E., Critchlow, S.E., Teo, S.H., Doherty, A.J., Priestley, A., Broughton, B., Kyse-la, B., Beamish, H., Plowman, N., Arlett, C.F., *et al.* (1999). Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. *Curr Biol* 9, 699-702.
- ♦ Rivera-Calzada, A., Maman, J.D., Spagnolo, L., Pearl, L.H., y Llorca, O. (2005). Three-dimensional structure and regulation of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). *Structure* 13, 243-255.
- ♦ Rivera-Calzada, A., Spagnolo, L., Pearl, L.H., y Llorca, O. (2007). Structural model of full-length human Ku70-Ku80 heterodimer and its recognition of DNA and DNA-PKcs. *EMBO Rep* 8, 56-62.
- ♦ Roberts, A.J., Numata, N., Walker, M.L., Kato, Y.S., Malkova, B., Kon, T., Ohkura, R., Arisaka, F., Knight, P.J., Sutoh, K., *et al.* (2009). AAA+ Ring and linker swing mechanism in the dynein motor. *Cell* 136, 485-495.
- ♦ Rohl, C.A., Strauss, C.E., Chivian, D., y Baker, D. (2004). Modeling structurally variable regions in homologous proteins with rosetta. *Proteins* 55, 656-677.
- ♦ Rosenthal, P.B., y Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* 333, 721-745.
- ♦ Rothkamm, K., Kruger, I., Thompson, L.H., y Lobrich, M. (2003). Pathways of DNA double-strand break repair during the mammalian cell cycle. *Mol Cell Biol* 23, 5706-5715.
- ♦ Russel, D., Lasker, K., Phillips, J., Schneidman-Duhovny, D., Velazquez-Muriel, J.A., y Sali, A. (2009). The structural dynamics of macromolecular processes. *Curr Opin Cell Biol* 21, 97-108.
- ♦ Sali, A., Glaeser, R., Earnest, T., y Baumeister, W. (2003). From words to literature in structural proteomics. *Nature* 422, 216-225.
- ♦ Scheres, S.H., Nunez-Ramirez, R., Gomez-Llorente, Y., San Martin, C., Eggermont, P.P., y Carazo, J.M. (2007). Modeling experimental image formation for likelihood-based classification of electron microscopy data. *Structure* 15, 1167-1177.

- ♦ Scheres, S.H., Valle, M., Nunez, R., Sorzano, C.O., Marabini, R., Herman, G.T., y Carazo, J.M. (2005). Maximum-likelihood multi-reference refinement for electron microscopy images. *J Mol Biol* 348, 139-149.
- ♦ Schuck, P., y Rossmann, P. (2000). Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* 54, 328-341.
- ♦ Segurado, M., y Tercero, J.A. (2009). The S-phase checkpoint: targeting the replication fork. *Biol Cell* 101, 617-627.
- ♦ Sekiguchi, J.M., y Ferguson, D.O. (2006). DNA double-strand break repair: a relentless hunt uncovers new prey. *Cell* 124, 260-262.
- ♦ Sengupta, S., y Harris, C.C. (2005). p53: traffic cop at the crossroads of DNA repair and recombination. *Nat Rev Mol Cell Biol* 6, 44-55.
- ♦ Shechter, D., Costanzo, V., y Gautier, J. (2004). Regulation of DNA replication by ATR: signaling in response to DNA intermediates. *DNA Repair (Amst)* 3, 901-908.
- ♦ Shiloh, Y. (2001). ATM and ATR: networking cellular responses to DNA damage. *Curr Opin Genet Dev* 11, 71-77.
- ♦ Shuman, S. (2009). DNA ligases: progress and prospects. *J Biol Chem* 284, 17365-17369.
- ♦ Shuman, S., y Schwer, B. (1995). RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases. *Mol Microbiol* 17, 405-410.
- ♦ Sibanda, B.L., Critchlow, S.E., Begun, J., Pei, X.Y., Jackson, S.P., Blundell, T.L., y Pellegrini, L. (2001). Crystal structure of an Xrcc4-DNA ligase IV complex. *Nat Struct Biol* 8, 1015-1019.
- ♦ Smith, J., Riballo, E., Kysela, B., Baldeyron, C., Manolis, K., Masson, C., Lieber, M.R., Papadopoulos, D., y Jeggo, P. (2003). Impact of DNA ligase IV on the fidelity of end joining in human cells. *Nucleic Acids Res* 31, 2157-2167.
- ♦ Sorzano, C.O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M., y Pascual-Montano, A. (2004). XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol* 148, 194-204.
- ♦ Spagnolo, L., Rivera-Calzada, A., Pearl, L.H., y Llorca, O. (2006). Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Mol Cell* 22, 511-519.
- ♦ Spahn, C.M., y Penczek, P.A. (2009). Exploring conformational modes of macromolecules.

lecular assemblies by multiparticle cryo-EM. *Curr Opin Struct Biol* 19, 623-631.

♦ Stucki, M., y Jackson, S.P. (2006). gammaH2AX and MDC1: anchoring the DNA-damage-response machinery to broken chromosomes. *DNA Repair (Amst)* 5, 534-543.

♦ Subramanya, H.S., Doherty, A.J., Ashford, S.R., y Wigley, D.B. (1996). Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell* 85, 607-615.

♦ Tsai, C.J., Kim, S.A., y Chu, G. (2007). Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc Natl Acad Sci U S A* 104, 7851-7856.

♦ Uhlmann, F., y Hopfner, K.P. (2006). Chromosome biology: the crux of the ring. *Curr Biol* 16, R102-105.

♦ van Attikum, H., y Gasser, S.M. (2009). Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol* 19, 207-217.

♦ van Gent, D.C., Hoeijmakers, J.H., y Kanaar, R. (2001). Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* 2, 196-206.

♦ van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., *et al.* (2000). Single-particle electron cryo-microscopy: towards atomic resolution. *Q Rev Biophys* 33, 307-369.

♦ van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., y Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J Struct Biol* 116, 17-24.

♦ van Heel, M., y Schatz, M. (2005). Fourier shell correlation threshold criteria. *J Struct Biol* 151, 250-262.

♦ Villa, E., Sengupta, J., Trabuco, L.G., LeBarron, J., Baxter, W.T., Shaikh, T.R., Grassucci, R.A., Nissen, P., Ehrenberg, M., Schulten, K., *et al.* (2009). Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc Natl Acad Sci U S A* 106, 1063-1068.

♦ Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., y Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189-1191.

♦ Weller, G.R., y Doherty, A.J. (2001). A family of DNA repair ligases in bacteria? *FEBS Lett* 505, 340-342.

♦ Weterings, E., y Chen, D.J. (2008). The endless tale of non-homologous end-joining. *Cell Res* 18, 114-124.

♦ Weterings, E., y van Gent, D.C. (2004). The mechanism of non-homologous end-



joining: a synopsis of synapsis. *DNA Repair (Amst)* 3, 1425-1435.

♦ Wu, P.Y., Frit, P., Meesala, S., Dauvillier, S., Modesti, M., Andres, S.N., Huang, Y., Sekiguchi, J., Calsou, P., Salles, B., *et al.* (2009). Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Mol Cell Biol* 29, 3163-3172.

♦ Yano, K., Morotomi-Yano, K., Wang, S.Y., Uematsu, N., Lee, K.J., Asaithamby, A., Weterings, E., y Chen, D.J. (2008). Ku recruits XLF to DNA double-strand breaks. *EMBO Rep* 9, 91-96.

♦ Yu, X., Jin, L., y Zhou, Z.H. (2008a). 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature* 453, 415-419.

♦ Yu, Y., Mahaney, B.L., Yano, K., Ye, R., Fang, S., Douglas, P., Chen, D.J., y Lees-Miller, S.P. (2008b). DNA-PK and ATM phosphorylation sites in XLF/Cernunnos are not required for repair of DNA double strand breaks. *DNA Repair (Amst)* 7, 1680-1692.

♦ Yu, Y., Wang, W., Ding, Q., Ye, R., Chen, D., Merkle, D., Schriemer, D., Meek, K., y Lees-Miller, S.P. (2003). DNA-PK phosphorylation sites in XRCC4 are not required for survival after radiation or for V(D)J recombination. *DNA Repair (Amst)* 2, 1239-1252.

♦ Zhang, J., Baker, M.L., Schroder, G.F., Douglas, N.R., Reissmann, S., Jakana, J., Dougherty, M., Fu, C.J., Levitt, M., Ludtke, S.J., *et al.* (2010). Mechanism of folding chamber closure in a group II chaperonin. *Nature* 463, 379-383.

♦ Zhang, X., Settembre, E., Xu, C., Dormitzer, P.R., Bellamy, R., Harrison, S.C., y Grigorieff, N. (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc Natl Acad Sci U S A* 105, 1867-1872.

♦ Zhou, B.B., y Elledge, S.J. (2000). The DNA damage response: putting checkpoints in perspective. *Nature* 408, 433-439.

♦ Zou, L., y Elledge, S.J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* 300, 1542-1548.





**PUBLICACIONES**





## Electron microscopy of Xrcc4 and the DNA ligase IV–Xrcc4 DNA repair complex

María A. Recuero-Checa<sup>a</sup>, Andrew S. Doré<sup>b</sup>, Ernesto Arias-Palomo<sup>a</sup>, Angel Rivera-Calzada<sup>a</sup>,  
Sjors H.W. Scheres<sup>c</sup>, Joseph D. Maman<sup>b,1</sup>, Laurence H. Pearl<sup>b,\*\*</sup>, Oscar Llorca<sup>a,\*</sup>

<sup>a</sup> Centro de Investigaciones Biológicas (CIB), Consejo Superior de Investigaciones Científicas (CSIC), Ramiro de Maeztu 9, 28040 Madrid, Spain

<sup>b</sup> Section of Structural Biology and Cancer Research UK DNA Repair Enzyme Research Group, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

<sup>c</sup> Centro Nacional de Biotecnología (CNB), Consejo Superior de Investigaciones Científicas (CSIC), Darwin 3, Cantoblanco, 28049 Madrid, Spain

### ARTICLE INFO

#### Article history:

Received 5 December 2008

Received in revised form

16 September 2009

Accepted 16 September 2009

#### Keywords:

Xrcc4

DNA ligase IV

Lig4

DNA repair

NHEJ

Electron microscopy

### ABSTRACT

The DNA ligase IV–Xrcc4 complex is responsible for the ligation of broken DNA ends in the non-homologous end-joining (NHEJ) pathway of DNA double strand break repair in mammals. Mutations in DNA ligase IV (Lig4) lead to immunodeficiency and radiosensitivity in humans. Only partial structural information for Lig4 and Xrcc4 is available, while the structure of the full-length proteins and their arrangement within the Lig4–Xrcc4 complex is unknown. The C-terminal domain of Xrcc4, whose structure has not been solved, contains phosphorylation sites for DNA-PKcs and is phylogenetically conserved, indicative of a regulatory role in NHEJ. Here, we have purified full length Xrcc4 and the Lig4–Xrcc4 complex, and analysed their structure by single-particle electron microscopy. The three-dimensional structure of Xrcc4 at a resolution of ~37 Å reveals that the C-terminus of Xrcc4 forms a dimeric globular domain connected to the N-terminus by a coiled-coil. The N- and C-terminal domains of Xrcc4 locate at opposite ends of an elongated molecule. The electron microscopy images of the Lig4–Xrcc4 complex were examined by two-dimensional image processing and a double-labelling strategy, identifying the site of the C-terminus of Xrcc4 and the catalytic core of Lig4 within the complex. The catalytic domains of Lig4 were found to be in the vicinity of the N-terminus of Xrcc4. We provide a first sight of the structural organization of the Lig4–Xrcc4 complex, which suggests that the BRCT domains could provide the link of the ligase to Xrcc4 while permitting some movements of the catalytic domains of Lig4. This arrangement may facilitate the ligation of diverse configurations of damaged DNA.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

DNA double-strand breaks (DSBs) are generated by genotoxic agents such as ionizing radiation and radiomimetic chemicals, by fork collapse during DNA replication, and as intermediates during physiological cellular processes like the V(D)J recombination that provides antibody diversity in developing lymphocytes. In mammalian cells that are not actively dividing, the predominant mechanism for repair of adventitious DSBs and the exclusive mechanism for breaks arising during V(D)J recombination is non-homologous end joining (NHEJ) [1–3]. Inadequacy or defects in these repair processes can lead to immunodeficiency, radiosensitivity and potentially to oncogenic transformation [3–6].

Genetic, biochemical and structural studies have characterized some of the proteins and mechanisms that participate in NHEJ [7–10]. Repair is performed by a multi-protein macromolecular complex assembled around the DNA-dependent protein kinase, a large enzyme comprising the ~470 kDa DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and the Ku70–Ku80 heterodimer. Initial recognition of the broken DNA ends is a function of the Ku70–Ku80 dimer. Crystal [11] and electron microscopy (EM) [12] structures of the Ku70–Ku80 dimer show that the protein can bind accessible DNA ends through a preformed ring in its structure. Ku70–Ku80 can then recruit DNA-PKcs whose DNA-dependent kinase activity becomes activated. The three-dimensional (3D) structure of the DNA-PKcs–Ku70–Ku80 complex assembled on DNA [13,14] suggests that the complex maintains broken DNA ends in proximity providing a platform for the recruitment of other proteins and enzymes required to complete the repair of the damage, including Artemis [15] and the Lig4–Xrcc4 complex [16,17]. A new NHEJ factor was recently identified by two groups and named XLF (Xrcc4-like factor) and Cernunnos [18,19]. XLF/Cernunnos is a 33 kDa protein with weak sequence homology to Xrcc4. Patients with defects in XLF/Cernunnos show clinical DNA

\* Corresponding author. Tel.: +34 91 837 3112x4446; fax: +34 91 536 0432.

\*\* Corresponding author. Tel.: +44 20 7153 5422; fax: +44 20 7153 5457.

E-mail addresses: [Laurence.Pearl@icr.ac.uk](mailto:Laurence.Pearl@icr.ac.uk) (L.H. Pearl), [ollorca@cib.csic.es](mailto:ollorca@cib.csic.es) (O. Llorca).

<sup>1</sup> Present address: Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK.